
JOINT TRAINING DOES NOT TRANSFER INFORMATION BETWEEN EEG AND IMAGE CLASSIFIERS

Anonymous authors

Paper under double-blind review

ABSTRACT

Caution is necessary with machine-learning methods, and especially computer-vision methods, to support brain processing claims from neuroimaging data. Recent papers propose (i) a joint-training process that does not use class information and (ii) a bidirectional transfer of (a) image information to an EEG classifier and (b) brain-activity information to an image classifier, such that the joint embedding includes the shared image and brain-activity information. These claims cannot be maintained: the training process is initialized with class information, and joint training with EEG degrades rather than improves the performance of the image encoder. Moreover, theoretical solutions exist that entail no transfer beyond class information in the joint embedding space.

1 INTRODUCTION

Recent papers (Palazzo et al., 2018; 2020a;b; 2021) propose a method for jointly training two neural networks, one to map images to encodings and the other to map EEG data from subjects viewing those images to encodings, so that the encodings for the images are similar to those from the EEG data. Specifically, they claim that

- I) *class labels are not used anywhere in the equation. This makes sure that the resulting embedding does not just associate class discriminative vectors to EEG and images, but tries to extract more comprehensive patterns that explain the relations between the two data modalities.* (Palazzo et al., 2021, § 3 ¶ 5) and
- II) *our multimodal approach learns a joint brain-visual embedding and finds similarities between brain representations and visual features* (Palazzo et al., 2021, § 1 ¶ 2).

They use the resulting encodings for EEG classification, image classification, saliency detection, and producing activation maps that decode brain representations. Central are the additional implicit claims

- III) that the joint embedding space encodes both information about the images and subjects' brain activity from viewing the images and
- IV) that the joint-training process bidirectionally transfers information from the image encoders to the EEG encoders, and information from the EEG encoders to the image encoders.

Other work (Li et al., 2021; Ahmed et al., 2021; 2022; Bharadwaj et al., 2023) questioned these claims due to confounds in the data (Spampinato et al., 2017) used by Palazzo et al. (2018; 2020a;b; 2021), which was collected with a block design with all stimuli of a given class presented to subjects in close proximity and the training and test sets containing samples from the same block. Since EEG data contains temporal drift, this drift is classified rather than stimulus-related brain activity. Li et al. (2021), Ahmed et al. (2021; 2022), and Bharadwaj et al. (2023) note that it is critical to remove the confound by breaking the correlation between temporal drift and stimulus class by randomizing stimulus presentation order. Li et al. (2021), Ahmed et al. (2021; 2022), and Bharadwaj et al. (2023) demonstrated that numerous classifiers, including the one used in Palazzo et al. (2018; 2020a;b; 2021), lose high classification accuracy seen on the data from Spampinato et al. (2017) used in Palazzo et al. (2018; 2020a;b; 2021), dropping to chance with nonconfounded data properly collected from randomized trials.

Here, we focus on a separate issue. We demonstrate that their method itself does not exhibit claims I–IV, whether applied to confounded or nonconfounded data. We contribute the novel observations

054 that independently refute the claims in Palazzo et al. (2018; 2020a;b; 2021) beyond Li et al. (2021),
055 Ahmed et al. (2021; 2022), and Bharadwaj et al. (2023):

- 056
057 A) Prior to joint training, the pretrained image encoders contain close to perfect class information,
058 and likely very little information other than class information. This calls claim I above into
059 question.
060 B) The models used as EEG and image encoders, and the loss function used to jointly train those
061 encoders, have the representational capacity to memorize one-hot EEG and image encodings
062 that minimize the loss function on the training set. These models can be found with separate
063 training. Thus it is unlikely that joint training transfers information from the image encoders to
064 the EEG encoders or vice versa, or includes anything other than class information. This calls
065 claims II–IV above into question.

066 We demonstrate that their method doesn’t support their claims, even when applied to data that reme-
067 dies the confound. Since their claims are contingent upon production of joint embeddings that in-
068 clude both image and brain-activity information, this calls all claims in Palazzo et al. (2018; 2020a;b;
069 2021) into question.

070 Beyond this, we further demonstrate

- 071 C) that joint training appears to (1) use the class information included in the pretrained image
072 encoders to train the EEG encoder as a classifier, but (2) otherwise degrades performance of the
073 image encoder as a classifier and
074 D) that while the EEG encoders so trained can perform above chance on confounded data, they
075 critically perform at chance on nonconfounded data.

076 Thus their joint-training method is ineffective.

077 Here, we limit our analyses and discussion to claims A–D, in so far as they refute claims I–IV. This
078 does not hinge on the confound reported in Li et al. (2021), Ahmed et al. (2021; 2022), and Bharad-
079 waj et al. (2023); in fact, we perform the same analyses both on confounded and nonconfounded
080 data, obtaining the same results. Our logical arguments all are based on the information content in
081 the encodings at various stages of various kinds of training, where information content is measured
082 by principal component analysis (PCA). It is expressly not based on absolute classification accu-
083 racy, *i.e.*, how well the classifiers perform. We are concerned with relative classification accuracy of
084 various reconstructions of the encodings with limited sets of components derived through PCA as a
085 way of assessing information content of the encodings.

086 Palazzo et al. (2018; 2020a;b; 2021) use the EEG and image encodings produced by jointly trained
087 EEG and image encoders to produce saliency and activation maps. The validity of these maps
088 hinges on the claim that these encodings contain both visual and brain-activity information. We
089 demonstrate this to be false, calling the validity of these maps into question. Li et al. (2021, § 5.4)
090 already questioned this validity both on methodological grounds and as a result of the confound.
091 Our results here offer additional independent reason to question the validity of these maps.

092 We make no claim that it is difficult or impossible to perform the central task attempted by Palazzo
093 et al. (2018; 2020a;b; 2021), namely to perform joint training of EEG and image encoders to transfer
094 image information to the EEG encoder and brain-activity information to the image encoder and yield
095 a joint embedding that contains both image and brain-activity information, without providing class
096 information. It might be possible to perform this task with a different method, or even with the
097 same method but with different data. Our sole claim is that their experiments and analyses do not
098 demonstrate that they have performed this central task as stated by claims I–IV made in Palazzo
099 et al. (2018; 2020a;b; 2021).

101 2 SIGNIFICANCE

102
103 Several independent lines of research have refuted a large body of flawed work (Spampinato et al.,
104 2017; Kavasidis et al., 2017; Palazzo et al., 2017; 2018; 2020a;b; 2021; 2024) along completely
105 different axes. Li et al. (2021) demonstrated that the dataset used (Spampinato et al., 2017), and the
106 methods used to collect that dataset, suffer from a temporal confound, correlating stimulus class with
107 experiment timing. Accuracy drops to chance when the confound is removed. Ahmed et al. (2021)
demonstrated that this holds even with a much larger dataset. Ahmed et al. (2022) demonstrated that

108 this holds for the additional classifiers used in Palazzo et al. (2018; 2020a;b; 2021). Bharadwaj et al.
109 (2023) demonstrated that this holds even when using supertrials.

110
111 Here we progress beyond prior demonstrations that the particular dataset (Spampinato et al., 2017)
112 is confounded. We offer the novel claim that the joint training regimen is also flawed. This is
113 significant because the same joint training regimen continues to be used (Bai et al., 2023; Zeng
114 et al., 2023b; Ahmadiéh et al., 2023; Lan et al., 2023; Du et al., 2023; Liu et al., 2023; Song et al.,
115 2023; Ye et al., 2024). None of the work that jointly trains EEG and image classifiers even attempts
116 to assess whether joint training is effective in information transfer. In most cases, we can't perform
117 the analysis because the published work does not contain sufficient information to do so, including
118 but not limited to stimulus timing and presentation order. Without this, one cannot even be sure that
119 the data is free from confounds.

120 This is further significant for the following reasons:

- 121 • Nearly one hundred papers (An & Cho, 2016; Spampinato et al., 2016; Ben Said et al.,
122 2017; Bozal Chaves, 2017; Kavasidis et al., 2017; Palazzo et al., 2017; Parekh et al., 2017;
123 Spampinato et al., 2017; Zhang et al., 2017; Du et al., 2018; Fares et al., 2018; Kumar et al.,
124 2018; Palazzo et al., 2018; Piplani et al., 2018; Tirupattur et al., 2018; Wang et al., 2018;
125 Zhang & Liu, 2018; Zhang et al., 2018; Zhong et al., 2018; Du et al., 2019; Hwang et al.,
126 2019; Jiang et al., 2019; Jiao et al., 2019; Long et al., 2019; Mukherjee et al., 2019; Uys,
127 2019; Wang et al., 2019; Cudlenco et al., 2020; Fares et al., 2020; Li et al., 2020; Palazzo
128 et al., 2020a;b; Wang et al., 2020; Zheng et al., 2020a;b; Palazzo et al., 2021; Zheng &
129 Chen, 2021; Ma et al., 2021; Mo et al., 2021; Jiang et al., 2021; Lee et al., 2021; Cavazza
130 et al., 2022; Khaleghi et al., 2022; Lee et al., 2022; Mishra et al., 2022; Mishra, 2022;
131 Scharnagl & Groth, 2022; Shimizu & Srinivasan, 2022; Ahmadiéh et al., 2023; Bai et al.,
132 2023; Du et al., 2023; Duan et al., 2023; Hasan & A, 2023; Imani et al., 2023; Lan et al.,
133 2023; Lee et al., 2023; Liu et al., 2023; Singh et al., 2023; Song et al., 2023; Wahengbam
134 et al., 2023; Zeng et al., 2023b;a; Fan et al., 2024; Ferrante et al., 2024a;b; Gou et al.,
135 2024; Lei et al., 2024; Liu et al., 2024a;b; Luvsansambuu et al., 2024; Mishra et al., 2024;
136 Mwata-Velu et al., 2024; Ngo et al., 2024; Palazzo et al., 2024; Pan et al., 2024; Qian et al.,
137 2024; Singh et al., 2024; Tang et al., 2024; de la Torre-Ortiz et al., 2024; Yang & Liu, 2024;
138 Ye et al., 2024; Zheng et al., 2024b;a; Zhu et al., 2024; Deng et al., 2025; Fares, 2025; Fu
139 et al., 2025; Lopez et al., 2025; Mehmood et al., 2025; Singh et al., 2025; Xiang et al.,
140 2025) draw flawed conclusions based on the confounded dataset from Spampinato et al.
141 (2017) and datasets suffering from the same confound.
- 142 • A number of new datasets have been collected with this same confounded protocol (Gou
143 et al., 2024; Pan et al., 2024; Zhu et al., 2024; Qian et al., 2024; Uys, 2019; Shimizu &
144 Srinivasan, 2022; Liu et al., 2024b; Wang et al., 2019; 2020; Ma et al., 2021; Cudlenco
145 et al., 2020; Zheng et al., 2024b; Cavazza et al., 2022; Luvsansambuu et al., 2024; Liu
146 et al., 2023; Bai et al., 2023; Parekh et al., 2017).
- 147 • A number of these have been publicly released and are used by others. For example, Singh
148 et al. (2023), Singh et al. (2024), and Lopez et al. (2025) use the dataset reported in Kumar
149 et al. (2018) and Duan et al. (2023), Singh et al. (2024), and Lopez et al. (2025) use the
150 dataset reported in Ma et al. (2021).
- 151 • This is further egregious because Palazzo et al. (2020b; 2024) continue to claim that their
152 dataset (Spampinato et al., 2017), and their results that were obtained with that dataset
153 (Spampinato et al., 2017; Kavasidis et al., 2017; Palazzo et al., 2017; 2018; 2020a;b; 2021;
154 2024), are valid, despite the refutations in Li et al. (2021), Ahmed et al. (2021; 2022), and
155 Bharadwaj et al. (2023), in part, because of claims I–IV in Palazzo et al. (2018; 2020a;b;
156 2021).
- 157 • This has been used to justify continued publication of a large and growing body of flawed
158 work based on confounded datasets (Cavazza et al., 2022; Khaleghi et al., 2022; Lee et al.,
159 2022; Mishra et al., 2022; Mishra, 2022; Scharnagl & Groth, 2022; Shimizu & Srinivasan,
160 2022; Ahmadiéh et al., 2023; Bai et al., 2023; Du et al., 2023; Duan et al., 2023; Hasan &
161 A, 2023; Imani et al., 2023; Lan et al., 2023; Lee et al., 2023; Liu et al., 2023; Singh et al.,
2023; Song et al., 2023; Wahengbam et al., 2023; Zeng et al., 2023b;a; Fan et al., 2024;
Ferrante et al., 2024a;b; Gou et al., 2024; Lei et al., 2024; Liu et al., 2024a;b; Luvsansam-
buu et al., 2024; Mishra et al., 2024; Mwata-Velu et al., 2024; Ngo et al., 2024; Palazzo
et al., 2024; Pan et al., 2024; Qian et al., 2024; Singh et al., 2024; Tang et al., 2024; de la

162 Torre-Ortiz et al., 2024; Yang & Liu, 2024; Ye et al., 2024; Zheng et al., 2024b;a; Zhu et al.,
163 2024; Deng et al., 2025; Fares, 2025; Fu et al., 2025; Lopez et al., 2025; Mehmood et al.,
164 2025; Singh et al., 2025; Xiang et al., 2025) even after the confound became known through
165 the work of Li et al. (2021), Ahmed et al. (2021; 2022), and Bharadwaj et al. (2023).

166 Current machine-learning conferences, and more generally, computer-science conferences and jour-
167 nals, are loathe to publish refutations. Observing this, Schaeffer et al. (2025) proposed that the field
168 of machine-learning establish a “refutations and critiques” track in prominent conferences. While
169 we applaud and support this proposal, the current lack of such a track should not be an impediment
170 to publishing refutations. Scientific journals in other fields have long done so, often resulting in
171 retraction of flawed work. Schaeffer et al. (2025) offer five example pieces of claimed flawed work
172 in machine learning. Each is an individual paper. These pale in comparison to the flaws we uncover
173 here: a systemic flaw of the entire peer review process across an entire field of inquiry, namely clas-
174 sification of stimulus image class from EEG recordings, that affects seventeen datasets and ninety
175 one papers. Moreover, none of the five examples in Schaeffer et al. (2025) are egregious; here the
176 authors of the flawed work continue to argue for its validity despite four refereed refutations and fifty
177 new flawed papers have been published subsequent to these four refereed refutations. This argues
178 for the need to make the community aware of the severity of the issue.

180 3 METHOD

181
182 We jointly train the EEG encoder (EEGChannelNet, Palazzo et al., 2018; 2020a;b; 2021) with
183 each of the same image encoders employed by Palazzo et al. (2018; 2020a;b; 2021): Inception
184 v3 (Szegedy et al., 2016), ResNet-101 (He et al., 2016), DenseNet-161 (Huang et al., 2017), and
185 AlexNet (Krizhevsky et al., 2012). Our analyses report results for all four image encoders, each an-
186 alyzed two ways: without pretraining (‘no pretraining’) and pretrained on the ILSVRC 2012 training
187 set (‘pretraining’). As we evaluate just the encodings, we do not postpend classifiers to these en-
188 coders. Thus these encoders produce 1000-element output vectors, not 40-element ones.

189
190 We train the encoders in two ways: (1) joint training as described in Palazzo et al. (2018; 2020a;b;
191 2021), with triplet loss, and (2) separate training individually with MSE loss against one-hot class
192 encodings. After separate training, we run a single epoch of joint training, without updating the
193 weights, just to compute the triplet loss. We report three kinds of results: ‘before’ is before any joint
194 or separate training, ‘after joint’ is after joint training, and ‘after separate’ is after separate training.

195 We repeat this analysis for three different EEG datasets: ‘Spampinato et al. (2017)’ refers to the data
196 from Spampinato et al. (2017) used by Palazzo et al. (2018; 2020a;b; 2021), ‘Li et al. (2021) block’
197 refers to subject 6 first image block run from Li et al. (2021), and ‘Li et al. (2021) randomized’
198 refers to subject 6 image rapid-event run from Li et al. (2021). ‘Spampinato et al. (2017)’ and ‘Li
199 et al. (2021) block’ are confounded; ‘Li et al. (2021) randomized’ is not.

200 We use the encoders, trained just on the training sets, to produce encodings on the validation and test
201 sets for all splits and pool the results. For the ‘Li et al. (2021) block’ and ‘Li et al. (2021) random-
202 ized’ data, the validation and test sets are a disjoint cover of the dataset so each sample has exactly
203 one encoding. Critically, the data from Spampinato et al. (2017) does not have this property. We
204 perform two analyses on these encodings. First, we perform principal-component analysis (PCA)
205 and reconstruct the encodings two ways: one with just the ‘top 40’ components and one with just
206 the ‘bottom 960’ components and report the fraction of variance in the encodings explained by the
207 top 40 components. Second, for both the raw encodings (‘original’) and the reconstructed encodings
208 (‘top 40’ and ‘bottom 960’), we compute the accuracy when using the encodings for classification,
209 without postpending any classifier (a 1000→40 FC layer followed by a softmax as done by Palazzo
210 et al. 2018; 2020a;b; 2021). Since the 40 classes employed by Palazzo et al. (2018; 2020a;b; 2021)
211 are a subset of the 1000 classes in ILSVRC 2012, and the image stimuli used by Spampinato et al.
212 (2017) to elicit EEG response are a subset of the training images in ILSVRC 2012, one can read
213 off the class label directly from the 1000-element encoding produced by either the EEG or image
214 encoders by choosing the index of the maximal element. We do this two ways. The first (‘1000
215 classes’) considers maximal elements outside of the 40 classes to be misclassification. The second
(‘40 classes’) ignores elements outside of the 40 classes and only computes the maximal element
among the 40 classes.

We also analyze the value of the loss function: for the triplet loss after both ‘joint’ and ‘separate’ training, on all four image encoders, both with and without ‘pretraining,’ and all three datasets, and also the MSE loss when separately training the ‘EEG’ and ‘image’ encoders. These losses are computed per sample and per element of the 1000-element encoding, averaged over samples and splits.

The appendix contains more details of our method.

4 RESULTS

Table 1 shows the fraction of variance explained by the top 40 principal components of the encodings produced for both modalities (‘EEG’ and ‘image’) for all image encoders and datasets, ‘before’ joint or separate training, ‘after joint’ training, both with and without ‘pretraining,’ and ‘after separate’ training.¹

- The top 40 principal components explain a large portion of the image-encoding variance ($\geq 63.9\%$) before training (column v; Table 5²). This together with Table 2 implies that the pretrained image encoders produce encodings that contain (primarily) class information on all data. This supports claim A and calls claim I into question.
- After joint training with pretraining on the ‘Spampinato et al. (2017)’ data, the top 40 principal components explain almost all ($\geq 94.9\%$) of the variance in the image encodings (column vi, for rows i–iv; Table 6). This implies that the image encoders jointly trained on the ‘Spampinato et al. (2017)’ data produce encodings with little more than class information.
- After separate training, the top 40 principal components explain almost all of the variance ($\geq 96.8\%$) in the image encodings (column viii; Table 7). This implies that the separately trained image encoders produce encodings with little more than class information on all data.
- After separate training, the top 40 principal components explain almost all of the variance ($\geq 96.8\%$) in both the EEG and image encodings on the ‘Spampinato et al. (2017)’ data (columns iv and viii, for rows i–iv; Table 8). This implies that the separately trained EEG and image encoders produce encodings with little more than class information on the ‘Spampinato et al. (2017)’ data.

Collectively this suggests that after either joint or separate training the image encodings contain class information and mostly class information. Likewise, after separate training on the ‘Spampinato et al. (2017)’ data, the EEG encodings contain primarily class information. This is not surprising, as has been noted, the ‘Spampinato et al. (2017)’ data is confounded. This supports claim B and calls claims II–IV into question.

Table 2 shows the accuracy of classifying the encodings of both modalities (‘EEG’ and ‘image’) for all image encoders and datasets, ‘before’ joint or separate training, ‘after joint’ training, both with and without ‘pretraining,’ and ‘after separate’ training, using the raw encodings (‘original’) or the encodings reconstructed from the ‘top 40’ or ‘bottom 960’ principal components, both when considering all ILSVRC 2012 classes (‘1000 classes’) and when only considering the ‘40 classes’ used by Palazzo et al. (2018; 2020a;b; 2021).

- Image classification accuracy is near perfect with all components with 40 classes ($\geq 95.5\%$; column i, for rows v–viii, xiii–xvi, and xxi–xxiv; Table 9) and very high with 1000 classes ($\geq 77.9\%$; column xiii, for rows v–viii, xiii–xvi, and xxi–xxiv; Table 10) before training. Likewise with just the top 40 principal components ($\geq 85.7\%$ and $\geq 72.2\%$; columns v and xvii, for rows v–viii, xiii–xvi, and xxi–xxiv; Table 11). This holds both for the ‘Spampinato et al. (2017)’ data and the ‘Li et al. (2021) block’ and ‘Li et al. (2021) randomized’ data. This demonstrates that they are giving class information implicitly to their training

¹Note that while the entries in Table 1 are percentages, they are not classification accuracies. Explained variance measures how close a reconstruction of the encodings using just the top 40 components is to the original encodings. High values indicate that the reconstruction is close to the original encodings because the bulk of the information in the encodings resides in the top 40 components.

²Here and throughout, tables numbered five and higher refer to those in the appendix. These contain variants of the five main tables with the indicated columns and rows highlighted in color.

Table 1: Explained variance (%) in the top 40 principal components of the encodings, ‘before’ joint or separate training, ‘after joint’ training, both with and without ‘pretraining,’ and ‘after separate’ training.

			EEG				image			
		before	after joint	after joint	after separate	before	after joint	after joint	after separate	
			pretraining	no pretraining		pretraining	no pretraining			
		i	ii	iii	iv	v	vi	vii	viii	
Spampinato et al. (2017)	Inception v3	i	44.7	37.2	15.3	99.5	64.1	94.9	98.9	100.0
	ResNet-101	ii	46.5	36.2	17.7	99.5	82.0	98.9	98.2	96.8
	DenseNet-161	iii	48.2	25.6	22.7	100.0	75.4	97.4	90.2	100.0
	AlexNet	iv	46.3	21.7	18.5	99.5	82.3	nan	100.0	99.7
Li et al. (2021) block	Inception v3	v	83.3	75.9	69.4	42.3	63.9	57.4	98.3	99.9
	ResNet-101	vi	85.4	79.8	72.8	41.9	81.9	99.5	99.0	99.7
	DenseNet-161	vii	84.0	79.3	76.3	42.9	75.3	69.4	94.3	99.5
	AlexNet	viii	84.4	78.9	73.2	40.0	82.2	78.9	99.9	98.9
Li et al. (2021) randomized	Inception v3	ix	62.1	40.5	38.3	19.3	63.9	60.2	98.1	99.9
	ResNet-101	x	61.9	52.2	43.6	19.7	81.9	99.1	98.8	99.8
	DenseNet-161	xi	63.5	48.5	46.5	19.6	75.3	70.8	93.9	99.7
	AlexNet	xii	63.6	49.5	41.3	19.2	82.2	86.9	99.9	98.8

by use of image encoders pretrained on ImageNet, supporting claim A, and calling claim I into question.

- Image classification accuracy is much lower ($\leq 46.0\%$) before training when discarding the top 40 principal components (columns ix and xxi, for rows v–viii, xiii–xvi, and xxi–xxiv; Table 12). This holds both for 40 classes and 1000 classes and both for the ‘Spampinato et al. (2017)’ data and the ‘Li et al. (2021) block’ and ‘Li et al. (2021) randomized’ data. Thus there is much less class information outside the top 40 principal components. This demonstrates that they are giving little more than class information implicitly to their training by use of image encoders pretrained on ImageNet, further supporting claim A and calling claim I into question.
- With only five exceptions, all in the bottom 960 principal components, image classification accuracy decreases after joint training with pretraining (compare columns ii, vi, x, xiv, xviii, and xxii to i, v, ix, xiii, xvii, and xxi, respectively, for rows v–viii, xiii–xvi, and xxi–xxiv; Table 13). This holds both for 40 classes and 1000 classes, both for the ‘Spampinato et al. (2017)’ data and the ‘Li et al. (2021) block’ and ‘Li et al. (2021) randomized’ data, and whether using all components or just the 40 principal components. This suggests that joint training is hurting, not helping, supporting claim C(2).
- With the exception of a small number of cases that are marginally above chance, EEG classification accuracy is at chance except after separate training on confounded data either on all components or the top 40 components (columns iv, viii, xvi, and xx, for rows i–iv and ix–xii; Table 14). It is also above chance after separate training on the ‘Li et al. (2021) block’ data in the bottom 960 components (columns xii and xxiv, for rows ix–xii; Table 15). This suggests that separate training is able to extract class from confounded data but not nonconfounded data and that joint training is not able to extract class from any data, supporting claims C(1) and D.
- While the image encoder can sometimes generalize (some of columns ii, iv, vi, viii, x, xii, xiv, xvi, xviii, xx, xxii, and xxiv, for rows v–viii, xiii–xvi, and xxi–xxiv, are above chance; Table 16) and the EEG encoder can sometimes generalize on confounded data with separate training (some of columns iv, viii, xii, xvi, xx, and xxiv, for rows i–iv and ix–xii, are above chance; Table 17), the EEG encoder cannot generalize on nonconfounded data with joint training with pretraining (all of columns ii, vi, x, xiv, xviii, and xxi, for rows xvii–xx, are at chance; Table 18). This supports claim D.
- Classification accuracy is at chance after joint training without pretraining (columns iii, vii, xi, xv, xix, and xxiii; Table 19). This holds both for EEG classification and image classification, both for 40 classes and 1000 classes, both for the ‘Spampinato et al. (2017)’ data and the ‘Li et al. (2021) block’ and ‘Li et al. (2021) randomized’ data, and whether using all components, just the 40 principal components, or the bottom 960 components. This suggests that their method completely breaks down when not provided with class information through pretraining, and further supports claim A and calls claim I into question.

The results in Table 2 exhibit some differences between different image encoders for each analysis (‘EEG’ vs. ‘image,’ ‘40 classes’ vs. ‘1000 classes,’ ‘original’ vs. ‘top 40’ vs. ‘bottom 960,’ and ‘before’ vs. ‘after joint’ vs. ‘after separate’). This is not surprising. Different image classifiers exhibit

324 different classification accuracy, especially when pretrained with different training regimens. It is
325 further not surprising that this difference manifests with different PCA reconstructions with different
326 components and also manifests when subjected to joint fine tuning with noisy EEG data or separate
327 fine tuning on the tiny subset of ImageNet used here, because fine tuning can break different models
328 in different ways. None of these differences impact our claims A–D or the refutation of claims I–IV
329 from Palazzo et al. (2018; 2020a;b; 2021).

330 Table 3 reports the per-sample triplet loss on the training set, after ‘joint’ training, both with and
331 without ‘pretraining,’ and ‘separate’ training, for all image encoders and datasets.

- 332 • With only two exceptions, separate training gets to a lower loss than joint training with
333 pretraining (compare columns vii, viii, and ix to i, ii, and iii, respectively; Table 20). This
334 critically suggests that there is a point within the representational-capacity space of their
335 model and loss function that achieves lower loss than achieved by their joint-training pro-
336 cedure. The joint-training procedure could have achieved it, it just didn’t. The fact that
337 point was achieved with separate training, indicates that the resulting EEG encodings do
338 not have any image information and the resulting image encodings do not have any brain-
339 activity information. This supports claim B and calls claims II–IV into question.
- 340 • With only three exceptions, joint training without pretraining gets to a lower loss than
341 with pretraining (compare columns iv, v, and vi to i, ii, and iii, respectively; Table 21).
342 Yet classification accuracy is at chance after joint training without pretraining (columns iii
343 vii, xi, xv, xix, and xxiii of Table 2; Table 19). This suggests that joint training without
344 pretraining can memorize the training set yet fail to generalize at all, and further supports
345 claim A and calls claim I into question.

346 Table 4 reports the per-sample MSE loss of the individual encoders of both modalities (‘EEG’ and
347 ‘image’) on the training set, after separate training, for all image encoders and datasets.

- 348 • Separate training against one-hot labels can get to very low MSE (all columns and rows),
349 indicating that the model has the representational capacity to memorize both the EEG and
350 image training data. Achieving this point in the representational-capacity space with sepa-
351 rate training and the encodings being one-hot implies that at this point, the EEG and image
352 encodings have nothing but class information. Since this point could have been achieved
353 with joint training, joint training could have produced a set of encodings that have nothing
354 but class information. This supports claim B and calls claims II–IV into question.

355 Collectively our results suggest that:

- 356 • Their statement (claim I) about not incorporating any class information in joint training is
357 false (our claim A).
- 358 • Their encoder models combined with their triplet loss function allow a point in the space
359 where the encodings on the training set are one-hot and thus cannot possibly contain any-
360 thing other than class information. This point has zero triplet loss and thus is the minimum.
361 It is achievable. Thus their framework can produce encodings with nothing but class infor-
362 mation.
- 363 • This point can be reached by an alternate separate-training regimen that clearly does not
364 transfer any information between the EEG and image encoders.

365 It is highly unlikely that the suboptimal EEG model that they produce by joint training has any image
366 information, the suboptimal image model that they produce by joint training has any brain-activity
367 information, and neither suboptimal model produced by joint training has anything other than class
368 information, even when run on nonconfounded data. This supports claim B and calls claims II–IV
369 into question.

371 5 CONCLUSION

372 We implore all future EEG image classification effort to release **raw data** that includes stimulus
373 timing and presentation order, not preprocessed data. We further implore all future effort to jointly
374 train EEG and image classifiers to employ the methods presented here to assess the effectiveness of
375 purported information transfer attributed to joint training, and not to assume such effectiveness due
376 to classification accuracy and image reconstruction.
377

378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431

Table 2: Original classification accuracy and classification accuracy after reconstruction (%), using just the encodings, ‘before’ joint or separate training, ‘after joint’ training, both with and without ‘pretraining,’ and ‘after separate’ training.

	40 classes				bottom 960				1000 classes				bottom 960						
	before	after separate	after joint	pretraining	before	after separate	after joint	pretraining	before	after separate	after joint	pretraining	before	after separate	after joint	pretraining			
Szepesvári et al. (2017)	i	2.6	2.8	3.0	2.6	2.2	2.2	2.2	ix	2.3	2.3	2.3	xxi	0.1	0.1	0.1	xxiv	0.0	
	ii	2.5	2.6	2.6	2.4	2.3	2.3	2.3	x	2.3	2.3	2.3	xxii	0.1	0.1	0.1	xxv	0.0	
	iii	2.3	2.4	2.4	2.4	2.4	2.4	2.4	xi	2.3	2.3	2.3	xxiii	0.1	0.1	0.1	xxvi	0.0	
	iv	2.3	2.4	2.4	2.6	2.6	2.6	2.6	xii	2.3	2.3	2.3	xxiv	0.1	0.1	0.1	xxvii	0.0	
	v	99.7	43.2	2.7	3.4	99.4	13.5	2.7	xiii	92.8	16.5	0.1	0.1	xxv	0.0	0.0	xxviii	0.0	
	vi	99.4	24.2	2.4	82.5	98.3	7.3	2.9	xiv	39.4	90.5	2.6	0.0	xxvi	0.0	0.0	xxix	0.0	
	vii	99.4	24.2	2.4	82.5	98.3	7.3	2.9	xv	90.5	2.6	0.0	0.0	xxvii	0.0	0.0	xxx	0.0	
	viii	95.8	2.6	2.7	18.7	85.8	2.6	2.7	xvi	78.4	0.0	0.3	0.3	xxviii	0.0	0.0	xxxi	0.0	
	ix	2.4	6.8	1.9	77.0	1.8	7.1	2.0	xvii	32.1	0.1	0.4	0.0	xxix	0.0	0.2	xxxii	0.1	
	x	1.6	3.5	2.2	76.6	1.1	4.2	2.5	xviii	28.1	0.0	0.1	0.1	xxx	0.0	0.0	xxxiii	0.1	
	xi	1.4	4.1	2.0	77.0	1.2	3.8	1.9	xix	25.9	0.1	0.1	0.0	xxxi	0.0	0.0	xxxiv	0.1	
	Li et al. (2021) block	xii	97.8	96.4	3.0	93.3	97.3	97.7	3.1	xx	93.0	75.1	0.1	0.1	xxxii	0.0	0.0	xxxv	0.0
xiii		99.3	75.0	2.4	11.0	98.5	68.5	2.7	xxi	3.7	90.5	56.4	0.0	0.1	0.2	0.0	xxxvi	0.0	
xiv		99.1	95.8	2.5	5.7	97.7	86.8	2.9	xxii	2.6	89.5	71.1	0.1	0.2	0.4	0.2	0.0	xxxvii	0.0
xv		95.5	60.9	2.5	85.7	50.9	2.4	2.4	xxiii	2.3	77.9	25.4	0.0	0.0	0.0	0.0	0.0	xxxviii	0.0
xvi		2.5	2.5	2.5	2.5	2.5	2.5	2.5	xxiv	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	xxxix	0.0
xvii		2.5	2.5	2.5	2.5	2.5	2.5	2.5	xxv	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	xxx	0.0
xviii		2.5	2.5	2.5	2.5	2.5	2.5	2.5	xxvi	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	xxxi	0.0
xix		2.5	2.5	2.5	2.5	2.5	2.5	2.5	xxvii	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	xxxii	0.0
xx		2.5	2.5	2.5	2.5	2.5	2.5	2.5	xxviii	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	xxxiii	0.0
xxi		2.5	2.5	2.5	2.5	2.5	2.5	2.5	xxix	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	xxxiv	0.0
xxii		2.5	2.5	2.5	2.5	2.5	2.5	2.5	xxx	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	xxxv	0.0
Li et al. (2021) randomized		xxiii	2.6	2.8	2.4	2.8	2.0	2.8	2.0	xxxi	1.9	0.0	0.1	0.1	xxxvii	0.0	0.2	0.0	xxxix
	xxiv	2.1	2.2	2.2	2.2	2.2	2.2	2.2	xxxii	1.5	0.0	0.1	0.1	xxxviii	0.0	0.1	0.1	xxxvi	0.1
	xxv	2.5	3.4	2.4	98.4	47.4	84.7	3.0	xxxiii	2.0	0.0	0.1	0.1	xxxix	0.0	0.1	0.1	xxxvii	0.1
	xxvi	99.3	99.3	2.1	11.1	98.4	58.0	2.1	xxxiv	2.1	2.5	99.5	0.0	0.0	0.0	0.0	0.0	xxxviii	0.0
	xxvii	99.1	99.1	2.1	11.1	97.7	36.0	2.5	xxxv	3.0	89.5	30.5	0.1	0.0	0.0	0.0	0.0	xxxix	0.0
	xxviii	95.5	65.6	2.1	51.1	97.7	36.0	2.5	xxxvi	2.3	3.0	43.9	0.1	0.0	0.0	0.0	0.0	xxx	0.0
	xxix	95.5	65.6	2.1	51.1	97.7	36.0	2.5	xxxvii	2.3	3.0	43.9	0.1	0.0	0.0	0.0	0.0	xxxvi	0.0
	xxx	95.5	65.6	2.1	51.1	97.7	36.0	2.5	xxxviii	2.3	3.0	43.9	0.1	0.0	0.0	0.0	0.0	xxxvii	0.0
	xxxi	95.5	65.6	2.1	51.1	97.7	36.0	2.5	xxxix	2.3	3.0	43.9	0.1	0.0	0.0	0.0	0.0	xxxviii	0.0
	xxxii	95.5	65.6	2.1	51.1	97.7	36.0	2.5	xxx	2.3	3.0	43.9	0.1	0.0	0.0	0.0	0.0	xxxviii	0.0
	xxxiii	95.5	65.6	2.1	51.1	97.7	36.0	2.5	xxxvi	2.3	3.0	43.9	0.1	0.0	0.0	0.0	0.0	xxxix	0.0

432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485

Table 3: Per-sample triplet loss on the training set, after joint training, both with and without ‘pre-training,’ and after separate training, averaged over samples and splits.

		joint						separate		
		pretraining			no pretraining					
		Spampinato et al. (2017)	Li et al. (2021) block	Li et al. (2021) randomized	Spampinato et al. (2017)	Li et al. (2021) block	Li et al. (2021) randomized	Spampinato et al. (2017)	Li et al. (2021) block	Li et al. (2021) randomized
Inception v3	i	1.303	1.098	0.932	1.074	0.938	0.919	1.150	1.027	1.104
ResNet-101	ii	1.155	0.925	1.019	0.963	1.006	1.080	0.885	0.936	0.884
DenseNet-161	iii	1.027	1.080	1.047	0.988	0.919	0.929	1.007	0.982	1.029
AlexNet	iv	1.417	1.142	0.981	1.136	1.055	1.034	0.925	0.956	0.955

Table 4: Per-sample MSE loss of the individual encoders on the training set, after separate training, averaged over samples and splits.

		EEG			image		
		Spampinato et al. (2017)	Li et al. (2021) block	Li et al. (2021) randomized	Spampinato et al. (2017)	Li et al. (2021) block	Li et al. (2021) randomized
Inception v3	i	0.001	0.001	0.001	0.002	0.010	0.010
ResNet-101	ii	0.001	0.001	0.001	0.001	0.002	0.002
DenseNet-161	iii	0.001	0.001	0.001	0.001	0.002	0.002
AlexNet	iv	0.001	0.001	0.001	0.001	0.001	0.001

Recent progress in computer vision and machine learning has been gauged largely by improvement in performance metrics of methods on datasets and results that ‘look good.’ Palazzo et al. (2018; 2020a;b; 2021) largely base their claims on similar criteria. While synthetic engineering methods can be evaluated by the utility of artifacts, analytic scientific claims about the underlying physical world can only be evaluated by finding hypotheses that resist falsification. We have falsified all of the claims of Palazzo et al. (2018; 2020a;b; 2021).

AUTHOR CONTRIBUTIONS

Removed for blind review.

ACKNOWLEDGMENTS

Removed for blind review.

ETHICS STATEMENT

This work debunks nearly one hundred published papers whose results are based on the same confound: a correlation between stimulus class and temporal drift. This confound has been found in eighteen available EEG datasets. Just as with an inconsistent set of axioms one can prove anything, a confounded dataset can be used to support any claim, even ones that are false or absurd. That is what many recent publications based on this confound do: things like generating high fidelity renderings of images, or even 3D CAD models of objects, from EEG recordings.

A research community, knowingly or unknowingly, has discovered that one can use confounded datasets to churn out a plethora of flawed results without reviewers noticing. They have also discovered that one can collect new confounded datasets to churn out even more flawed results without reviewers noticing. The temptation to do this is so strong that the community continues to do so four years after details of the confound were published.

It is conceivable that the flaws in these datasets may be a driving factor behind their frequent reuse. When a dataset is severely confounded, it becomes relatively easy to achieve an extremely high accuracy, which can in turn be used to support sensational claims, and ultimately directs further attention to the dataset. In business, this phenomenon is referred to as “the bad money drives out the good money.”

More prominent exposure of these flawed methods and consequent false results will allow resources wasted on continued use of these confounded datasets and flawed methods to be reallocated. The debunked work also causes direct ongoing harm:

- grant proposals can be rejected due to preliminary results not being competitive with results demonstrating falsely-inflated performance based on confounded data or faulty methods;
- manuscripts can be rejected for the same reason;
- grants can be awarded based on false pretenses
- manuscripts can be accepted for the same reason;
- degrees can be awarded for the same reason;

-
- 486
- resources can be wasted attempting to replicate the debunked results;
 - resources can be wasted having people read and review flawed papers, and learn flawed methods; and
 - because the debunked work relates to brain-computer interfaces—whose primary application is helping people with disabilities (e.g., paralysis) interact with the world—the harm caused is not merely scientific but also medical, with disproportionate impact on people with disabilities.
- 487
488
489
490
491
492

493 REPRODUCIBILITY STATEMENT

494 The raw data that produced these results is available at <http://dx.doi.org/10.21227/x2gf-5324>. Our code, which will be released upon publication, is built on top of the code in <http://dx.doi.org/10.21227/x2gf-5324>.

495
496
497
498

499 REFERENCES

- 500
- 501 Hajar Ahmadieh, Farnaz Gassemi, and Mohammad Hasan Moradi. A hybrid deep learning frame-
502 work for automated visual image classification using EEG signals. *Neural Computing and Appli-*
503 *cations*, pp. 1–17, 2023.
- 504
- 505 Hamad Ahmed, Ronnie B. Wilbur, Hari M. Bharadwaj, and Jeffrey Mark Siskind. Object classifica-
506 tion from randomized EEG trials. In *Computer Vision and Pattern Recognition*, pp. 3845–3854,
507 2021.
- 508
- 509 Hamad Ahmed, Ronnie B. Wilbur, Hari M. Bharadwaj, and Jeffrey Mark Siskind. Confounds in the
510 data—Comments on “Decoding brain representations by multimodal learning of neural activity
511 and visual features”. *Transactions on Pattern Analysis and Machine Intelligence*, 44(12):9217–
512 9220, 2022.
- 513
- 514 Jinwon An and Sungzoon Cho. Hand motion identification of grasp-and-lift task from electroen-
515 cephalography recordings using recurrent neural networks. In *International Conference on Big
Data and Smart Computing*, pp. 427–429, 2016.
- 516
- 517 Yunpeng Bai, Xintao Wang, Yan-pei Cao, Yixiao Ge, Chun Yuan, and Ying Shan. Dreamdiffusion:
518 Generating high-quality images from brain EEG signals. *arXiv*, 2306.16934, 2023.
- 519
- 520 Ahmed Ben Said, Amr Mohamed, Tarek Elfouly, Khaled Harras, and Z. Jane Wang. Multimodal
521 deep learning approach for joint EEG-EMG data compression and classification. In *Wireless
Communications and Networking Conference*, 2017.
- 522
- 523 Hari M. Bharadwaj, Ronnie B. Wilbur, and Jeffrey Mark Siskind. Still an ineffective method with
524 supertrials/ERPs—Comments on “Decoding brain representations by multimodal learning of neu-
525 ral activity and visual features”. *Transactions on Pattern Analysis and Machine Intelligence*, 45
526 (11):14052–14054, 2023.
- 527
- 528 Alberto Bozal Chaves. Personalized image classification from EEG signals using deep learning.
B.S. thesis, Universitat Politècnica de Catalunya, 2017.
- 529
- 530 Jacopo Cavazza, Waqar Ahmed, Riccardo Volpi, Pietro Morerio, Francesco Bossi, Cesco Willemse,
531 Agnieszka Wykowska, and Vittorio Murino. Understanding action concepts from videos and
532 brain activity through subjects’ consensus. *Scientific Reports*, 12(1):19073, 2022.
- 533
- 534 Nicolae Cudlenco, Nirvana Popescu, and Marius Leordeanu. Reading into the mind’s eye: Boosting
automatic visual recognition with EEG signals. *Neurocomputing*, 386:281–292, 2020.
- 535
- 536 Carlos de la Torre-Ortiz, Michiel M. Spapé, Niklas Ravaja, and Tuukka Ruotsalo. Cross-subject
537 EEG feedback for implicit image generation. *Transactions on Cybernetics*, 54(10):6105–6117,
538 2024.
- 539
- Xia Deng, Shen Chen, Jiale Zhou, and Lei Li. Mind2matter: Creating 3D models from EEG signals.
arXiv, 2504.11936, 2025.

-
- 540 Changde Du, Changying Du, and Huiguang He. Doubly semi-supervised multimodal adversarial
541 learning for classification, generation and retrieval. In *International Conference on Multimedia*,
542 pp. 13–18, 2019.
- 543 Changde Du, Kaicheng Fu, Jinpeng Li, and Huiguang He. Decoding visual neural representations
544 by multimodal learning of brain-visual-linguistic features. *Transactions on Pattern Analysis and*
545 *Machine Intelligence*, 45(9):10760–10777, 2023.
- 546
- 547 Changying Du, Changde Du, Xingyu Xie, Chen Zhang, and Hao Wang. Multi-view adversarially
548 learned inference for cross-domain joint distribution matching. In *International Conference on*
549 *Knowledge Discovery & Data Mining*, pp. 1348–1357, 2018.
- 550
- 551 Yiping Duan, Shuzhan Hu, Xin Ma, and Xiaoming Tao. Multi-class image generation from EEG
552 features with conditional generative adversarial networks. In *International Conference on Wire-*
553 *less Communications and Signal Processing*, pp. 534–539, 2023.
- 554 Xiaoya Fan, Yuntao Liu, and Zhong Wang. Electroencephalogram helps few-shot learning. In
555 *International Conference on Acoustics, Speech and Signal Processing*, pp. 8015–8019, 2024.
- 556
- 557 Ahmed Fares. A novel spatiotemporal framework for EEG-based visual image classification through
558 signal disambiguation. *Applied System Innovation*, 8(5):121, 2025.
- 559 Ahmed Fares, Shenghua Zhong, and Jianmin Jiang. Region level bi-directional deep learning frame-
560 work for EEG-based image classification. In *International Conference on Bioinformatics and*
561 *Biomedicine*, pp. 368–373, 2018.
- 562 Ahmed Fares, Sheng-hua Zhong, and Jianmin Jiang. Brain-media: A dual conditioned and lateral-
563 ization supported GAN (DCLS-GAN) towards visualization of image-evoked brain activities. In
564 *International Conference on Multimedia*, pp. 1764–1772, 2020.
- 565
- 566 Matteo Ferrante, Tommaso Boccato, Stefano Bargione, and Nicola Toschi. Decoding EEG signals
567 of visual brain representations with a CLIP based knowledge distillation. In *Learning from Time*
568 *Series For Health*, 2024a.
- 569 Matteo Ferrante, Tommaso Boccato, Stefano Bargione, and Nicola Toschi. Decoding visual brain
570 representations from electroencephalography through knowledge distillation and latent diffusion
571 models. *Computers in Biology and Medicine*, 178:108701, 2024b.
- 572
- 573 Honghao Fu, Hao Wang, Jing Jih Chin, and Zhiqi Shen. BrainVis: Exploring the bridge between
574 brain and visual signals via image reconstruction. In *International Conference on Acoustics,*
575 *Speech and Signal Processing*, pp. 1–5, 2025.
- 576
- 577 Mingyu Gou, Ying-Jie Zhang, Ren-Jie Dai, Hao-Long Yin, Tianzhen Chen, Fei Cheng, Bao-Liang
578 Lu, Jiang Du, and Wei-Long Zheng. Addressing temporal and auditory factors in meditative EEG
579 with self-supervised learning. In *International Conference on Bioinformatics and Biomedicine*,
pp. 1954–1959, 2024.
- 580
- 581 Haitham S Hasan and Al-Sharqi Mais A. EEG-based image classification using an efficient geo-
582 metric deep network based on functional connectivity. *Periodicals of Engineering and Natural*
583 *Sciences*, 11(1):208–215, 2023.
- 584
- 585 Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recog-
586 nition. In *Computer Vision and Pattern Recognition*, pp. 770–778, 2016.
- 587
- 588 Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected
589 convolutional networks. In *Computer Vision and Pattern Recognition*, pp. 4700–4708, 2017.
- 590
- 591 Sunhee Hwang, Kibeom Hong, Guiyoung Son, and Hyeran Byun. EZSL-GAN: EEG-based zero-
592 shot learning approach using a generative adversarial network. In *International Winter Conference*
593 *on Brain-Computer Interface*, pp. 1–4, 2019.
- 594
- 595 Zahra Imani, Mehdi Ezoji, and Timothée Masquelier. Brain-guided manifold transferring to improve
596 the performance of spiking neural networks in image classification. *Journal of Computational*
597 *Neuroscience*, 51(4):475–490, 2023.

-
- 594 Jianmin Jiang, Ahmed Fares, and Sheng-Hua Zhong. A context-supported deep learning framework
595 for multimodal brain imaging classification. *Transactions on Human-Machine Systems*, 49(6):
596 611–622, 2019.
- 597
598 Jianmin Jiang, Ahmed Fares, and Sheng-Hua Zhong. A brain-media deep framework towards seeing
599 imaginations inside brains. *Transactions on Multimedia*, 23:1454–1465, 2021.
- 600
601 Zhicheng Jiao, Haoxuan You, Fan Yang, Xin Li, Han Zhang, and Dinggang Shen. Decoding EEG by
602 visual-guided deep neural networks. In *International Joint Conference on Artificial Intelligence*,
603 2019.
- 604
605 Isaak Kavasidis, Simone Palazzo, Concetto Spampinato, Daniela Giordano, and Mubarak Shah.
606 Brain2Image: Converting brain signals into images. In *International Conference on Multimedia*,
607 pp. 1809–1817, 2017.
- 608
609 Nastaran Khaleghi, Tohid Yousefi Rezaei, Soosan Beheshti, Saeed Meshgini, Sobhan Sheykhivand,
610 and Sebelan Danishvar. Visual saliency and image reconstruction from EEG signals via an ef-
611 fective geometric deep network-based generative adversarial network. *Electronics*, 11(21):3637,
612 2022.
- 613
614 Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet classification with deep convo-
615 lutional neural networks. In *Advances in Neural Information Processing Systems*, pp. 1097–1105,
616 2012.
- 617
618 Pradeep Kumar, Rajkumar Saini, Partha Pratim Roy, Pawan Kumar Sahu, and Debi Prosad Dogra.
619 Envisioned speech recognition using EEG sensors. *Personal and Ubiquitous Computing*, 22(1):
620 185–199, 2018.
- 621
622 Yu-Ting Lan, Kan Ren, Yansen Wang, Wei-Long Zheng, Dongsheng Li, Bao-Liang Lu, and Lili Qiu.
623 Seeing through the brain: image reconstruction of visual perception from human brain signals.
624 *arXiv*, 2308.02510, 2023.
- 625
626 Pilhyeon Lee, Sunhee Hwang, Seogkyu Jeon, and Hyeran Byun. Subject adaptive EEG-based visual
627 recognition. In *Asian Conference on Pattern Recognition*, pp. 322–334, 2021.
- 628
629 Pilhyeon Lee, Sunhee Hwang, Jewook Lee, Minjung Shin, Seogkyu Jeon, and Hyeran Byun. Inter-
630 subject contrastive learning for subject adaptive EEG-based visual recognition. In *International
631 Winter Conference on Brain-Computer Interface*, pp. 1–6, 2022.
- 632
633 Pilhyeon Lee, Seogkyu Jeon, Sunhee Hwang, Minjung Shin, and Hyeran Byun. Source-free sub-
634 ject adaptation for EEG-based visual recognition. In *International Winter Conference on Brain-
635 Computer Interface*, pp. 1–6, 2023.
- 636
637 Weixian Lei, Yixiao Ge, Kun Yi, Jianfeng Zhang, Difei Gao, Dylan Sun, Yuying Ge, Ying Shan,
638 and Mike Zheng Shou. VIT-LENS: Towards omni-modal representations. In *Computer Vision
639 and Pattern Recognition*, pp. 26637–26647, 2024.
- 640
641 Dan Li, Changde Du, and Huiguang He. Semi-supervised cross-modal image generation with gener-
642 ative adversarial networks. *Pattern Recognition*, 100, 2020.
- 643
644 Ren Li, Jared S. Johansen, Hamad Ahmed, Thomas V. Ilyevsky, Ronnie B. Wilbur, Hari M. Bharad-
645 waj, and Jeffrey Mark Siskind. The perils and pitfalls of block design for EEG classification
646 experiments. *Transactions on Pattern Analysis and Machine Intelligence*, 43(1):316–333, 2021.
- 647
648 Dongjun Liu, Weichen Dai, Hangkui Zhang, Xuanyu Jin, Jianting Cao, and Wanzeng Kong. Brain-
649 machine coupled learning method for facial emotion recognition. *Transactions on Pattern Anal-
650 ysis and Machine Intelligence*, 45(9):10703–10717, 2023.
- 651
652 Dongjun Liu, Jin Cui, Zeyu Pan, Hangkui Zhang, Jianting Cao, and Wanzeng Kong. Machine to
653 brain: facial expression recognition using brain machine generative adversarial networks. *Cogni-
654 tive Neurodynamics*, 18(3):863–875, 2024a.

-
- 648 Xuan-Hao Liu, Yan-Kai Liu, Yansen Wang, Kan Ren, Hanwen Shi, Zilong Wang, Dongsheng Li,
649 Bao-Liang Lu, and Wei-Long Zheng. EEG2video: Towards decoding dynamic visual perception
650 from EEG signals. In *Advances in Neural Information Processing Systems*, pp. 72245–72273,
651 2024b.
- 652 Yanfang Long, Wanzeng Kong, Xuanyu Jin, Jili Shang, and Can Yang. Visualizing emotional states:
653 A method based on human brain activity. In *Human Brain and Artificial Intelligence*, pp. 248–
654 258, 2019.
- 656 Eleonora Lopez, Luigi Sigillo, Federica Colonnese, Massimo Panella, and Danilo Comminiello.
657 Guess what I think: Streamlined EEG-to-image generation with latent diffusion models. In *Inter-
658 national Conference on Acoustics, Speech and Signal Processing*, pp. 1–5, 2025.
- 660 Uurtsaikh Luvsansambuu, Tengis Tserendondog, Munkbayar Bat-Erdene, and Batmunkh Amar. A
661 deep learning model for classifying the thoughts of multiple individuals based on visual event. In
662 *International Conference on Electrical, Computer and Energy Technologies*, pp. 1–6, 2024.
- 663 Xin Ma, Yiping Duan, Shuzhan Hu, Xiaoming Tao, and Ning Ge. EEG based visual classification
664 with multi-feature joint learning. In *International Conference on Image Processing*, pp. 264–268,
665 2021.
- 667 Tariq Mehmood, Hamza Ahmad, Muhammad Haroon Shakeel, and Murtaza Taj. CATVis: Context-
668 aware thought visualization. *arXiv*, 2507.11522, 2025.
- 669 Abhijit Mishra, Shreya Shukla, Jose Torres, Jacek Gwizdka, and Shounak Roychowdhury.
670 Thought2text: Text generation from EEG signal using large language models (llms). *arXiv*,
671 2410.07507, 2024.
- 673 Alankrit Mishra. Enhancing machine vision using human cognition from EEG analysis. Master’s
674 thesis, Lakehead University, 2022.
- 675 Alankrit Mishra, Nikhil Raj, and Garima Bajwa. EEG-based image feature extraction for visual
676 classification using deep learning. In *International Conference on Intelligent Data Science Tech-
677 nologies and Applications*, pp. 181–188, 2022.
- 679 Liangyan Mo, Yuhan Wang, Wenhui Zhou, Xingfa Shen, and Wanzeng Kong. A bi-LSTM based
680 network with attention mechanism for EEG visual classification. In *International Conference on
681 Unmanned Systems*, pp. 858–863, 2021.
- 683 Pranay Mukherjee, Abhirup Das, Ayan Kumar Bhunia, and Partha Pratim Roy. Cogni-Net: Cog-
684 nitive feature learning through deep visual perception. In *International Conference on Image
685 Processing*, pp. 4539–4543, 2019.
- 686 Tat’y Mwata-Velu, Erik Zamora, Juan Irving Vasquez-Gomez, Jose Ruiz-Pinales, and Humberto
687 Sossa. Multiclass classification of visual electroencephalogram based on channel selection, min-
688 imum norm estimation algorithm, and deep network architectures. *Sensors*, 24(12):3968, 2024.
- 690 Huyen Ngo, Khoi Do, Duong Nguyen, Viet Dung Nguyen, and Lan Dang. How homogenizing the
691 channel-wise magnitude can enhance EEG classification model? *arXiv*, 2407.20247, 2024.
- 692 Simone Palazzo, Concetto Spampinato, Isaak Kavasidis, Daniela Giordano, and Mubarak Shah.
693 Generative adversarial networks conditioned by brain signals. In *International Conference on
694 Computer Vision*, pp. 3410–3418, 2017.
- 696 Simone Palazzo, Concetto Spampinato, Isaak Kavasidis, Daniela Giordano, and Mubarak Shah.
697 Decoding brain representations by multimodal learning of neural activity and visual features.
698 *arXiv*, 1820.10974v1, 2018.
- 700 Simone Palazzo, Francesco Rundo, Sebastiano Battiato, Daniela Giordano, and Concetto Spamp-
701 inato. Visual saliency detection guided by neural signals. In *International Conference on Auto-
matic Face and Gesture Recognition*, pp. 434–440, 2020a.

-
- 702 Simone Palazzo, Concetto Spampinato, Joseph Schmidt, Isaak Kavasidis, Daniela Giordano, and
703 Mubarak Shah. Correct block-design experiments mitigate temporal correlation bias in EEG
704 classification. *arXiv*, 2012.03849, 2020b.
- 705
706 Simone Palazzo, Concetto Spampinato, Isaak Kavasidis, Daniela Giordano, Joseph Schmidt, and
707 Mubarak Shah. Decoding brain representations by multimodal learning of neural activity and
708 visual features. *Transactions on Pattern Analysis and Machine Intelligence*, 43(11):3833–3849,
709 2021.
- 710 Simone Palazzo, Concetto Spampinato, Isaak Kavasidis, Daniela Giordano, Joseph Schmidt, and
711 Mubarak Shah. Rebuttal to “Comments on ‘Decoding brain representations by multimodal learn-
712 ing of neural activity and visual features’”. *Transactions on Pattern Analysis and Machine Intel-
713 ligence*, 46(12):11540–11542, 2024.
- 714
715 Hongguang Pan, Zhuoyi Li, Yunpeng Fu, Xuebin Qin, and Jianchen Hu. Reconstructing visual stim-
716 ulus representation from EEG signals based on deep visual representation model. *Transactions
717 on Human-Machine Systems*, 54(6):711–722, 2024.
- 718
719 Viral Parekh, Ramanathan Subramanian, Dipanjan Roy, and CV Jawahar. An EEG-based image
720 annotation system. In *National Conference on Computer Vision, Pattern Recognition, Image
721 Processing, and Graphics*, pp. 303–313, 2017.
- 722
723 Tanya Piplani, Nick Merrill, and John Chuang. Faking it, making it: Fooling and improving brain-
724 based authentication with generative adversarial networks. In *International Conference on Bio-
725 metrics Theory, Applications and Systems*, 2018.
- 726
727 Dongguan Qian, Hong Zeng, Wenjie Cheng, Yu Liu, Taha Bikki, and Jianjiang Pan. NeuroDM:
728 Decoding and visualizing human brain activity with EEG-guided diffusion model. *Computer
729 Methods and Programs in Biomedicine*, 251:108213, 2024.
- 730
731 Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng
732 Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei.
733 ImageNet large scale visual recognition challenge. *arXiv*, 1409.0575, 2014.
- 734
735 Rylan Schaeffer, Joshua Kazdan, Yegor Denisov-Blanch, Brando Miranda, Matthias Gerstgrasser,
736 Susan Zhang, Andreas Haupt, Isha Gupta, Elyas Obbad, Jesse Dodge, Jessica Zosa Forde,
737 Francesco Orabona, Sanmi Koyejo, and David Donoho. Position: Machine learning conferences
738 should establish a “refutations and critiques” track. *arXiv*, 2506.19882, 2025.
- 739
740 Bastian Scharnagl and Christian Groth. Evaluation of different deep learning approaches for EEG
741 classification. In *International Conference on Artificial Intelligence for Industries*, pp. 42–47,
742 2022.
- 743
744 Hirokatsu Shimizu and Ramesh Srinivasan. Improving classification and reconstruction of imagined
745 images from EEG signals. *Plos one*, 17(9):e0274847, 2022.
- 746
747 Prajwal Singh, Pankaj Pandey, Krishna Miyapuram, and Shanmuganathan Raman. EEG2IMAGE:
748 Image reconstruction from EEG brain signals. In *International Conference on Acoustics, Speech
749 and Signal Processing*, pp. 1–5, 2023.
- 750
751 Prajwal Singh, Dwip Dalal, Gautam Vashishtha, Krishna Miyapuram, and Shanmuganathan Raman.
752 Learning robust deep visual representations from EEG brain recordings. In *Winter Conference on
753 Applications of Computer Vision*, pp. 7538–7547, 2024.
- 754
755 Pushapdeep Singh, Jyoti Nigam, Medicherla Vamsi Krishna, Arnav Bhavsar, and Aditya Nigam.
EAD: An EEG adapter for automated classification. *arXiv*, 2505.23107, 2025.
- 756
757 Yonghao Song, Bingchuan Liu, Xiang Li, Nanlin Shi, Yijun Wang, and Xiaorong Gao. Decoding
758 natural images from EEG for object recognition. *arXiv*, 2308.13234, 2023.
- 759
760 Concetto Spampinato, Simone Palazzo, Isaak Kavasidis, Daniela Giordano, Mubarak Shah, and
761 Nasim Souly. Deep learning human mind for automated visual classification. *arXiv*, 1609.00344,
762 2016.

-
- 756 Concetto Spampinato, Simone Palazzo, Isaak Kavasidis, Daniela Giordano, Nasim Souly, and
757 Mubarak Shah. Deep learning human mind for automated visual classification. In *Computer*
758 *Vision and Pattern Recognition*, pp. 6809–6817, 2017.
- 759
760 Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking
761 the Inception architecture for computer vision. In *Computer Vision and Pattern Recognition*, pp.
762 2818–2826, 2016.
- 763 Jijia Tang, Yutao Yang, Qibin Zhao, Yu Ding, Jianhai Zhang, Yang Song, and Wanzeng Kong.
764 Visual-guided dual-spatial interaction network for fine-grained brain semantic decoding. *Trans-*
765 *actions on Instrumentation and Measurement*, 73:1–14, 2024.
- 766
767 Praveen Tirupattur, Yogesh Singh Rawat, Concetto Spampinato, and Mubarak Shah. ThoughtViz:
768 Visualizing human thoughts using generative adversarial network. In *International Conference*
769 *on Multimedia*, pp. 950–958, 2018.
- 770 Pieter Johannes Uys. Image classification from EEG brain signals using machine learning and deep
771 learning techniques. Master’s thesis, Stellenbosch University, 2019.
- 772
773 Kanan Wahengbam, Kshetrimayum Linthoinganbi Devi, and Aheibam Dinamani Singh. Fortifying
774 brain signals for robust interpretation. *Transactions on Network Science and Engineering*, 10(2):
775 742–753, 2023.
- 776 Fang Wang, Sheng Hua Zhong, Jianfeng Peng, Jianmin Jiang, and Yan Liu. Data augmentation
777 for EEG-based emotion recognition with deep convolutional neural networks. *Lecture Notes in*
778 *Computer Science*, 10705:82–93, 2018.
- 779
780 Pan Wang, Danlin Peng, Ling Li, Liuqing Chen, Chao Wu, Xiaoyi Wang, Peter Childs, and Yike
781 Guo. Human-in-the-loop design with machine learning. In *International Conference on Engi-*
782 *neering Design*, pp. 2577–2586, 2019.
- 783 Pan Wang, Shuo Wang, Danlin Peng, Liuqing Chen, Chao Wu, Zhen Wei, Peter Childs, Yike Guo,
784 and Ling Li. Neurocognition-inspired design with machine learning. *Design science*, 6:e33, 2020.
- 785
786 Xin Xiang, Wenhui Zhou, and Guojun Dai. Electroencephalography-driven three-dimensional ob-
787 ject decoding with multi-view perception diffusion. *Engineering Applications of Artificial Intel-*
788 *ligence*, 156:111180, 2025.
- 789 Guangyu Yang and Jinguo Liu. A new framework combining diffusion models and the convolution
790 classifier for generating images from EEG signals. *Brain Sciences*, 14(5):478, 2024.
- 791
792 Zesheng Ye, Lina Yao, Yu Zhang, and Sylvia Gustin. Self-supervised cross-modal visual retrieval
793 from brain activities. *Pattern Recognition*, 145:109915, 2024.
- 794
795 Hong Zeng, Nianzhang Xia, Dongguan Qian, Motonobu Hattori, Chu Wang, and Wanzeng Kong.
796 DM-RE2I: A framework based on diffusion model for the reconstruction from EEG to image.
797 *Biomedical Signal Processing and Control*, 86:105125, 2023a.
- 798
799 Hong Zeng, Nianzhang Xia, Ming Tao, Deng Pan, Haohao Zheng, Chu Wang, Feifan Xu, Wael Za-
800 karia, and Guojun Dai. DCAE: A dual conditional autoencoder framework for the reconstruction
801 from EEG into image. *Biomedical Signal Processing and Control*, 81:104440, 2023b.
- 802
803 Wenxiang Zhang and Qingshan Liu. Using the center loss function to improve deep learning per-
804 formance for EEG signal classification. In *International Conference on Advanced Computational*
805 *Intelligence*, pp. 578–582, 2018.
- 806
807 X. Zhang, L. Yao, Q. Z. Sheng, S. S. Kanhere, T. Gu, and D. Zhang. Converting your thoughts to
808 texts: Enabling brain typing via deep feature learning of EEG signals. In *International Conference*
809 *on Pervasive Computing and Communications*, 2018.
- 808
809 Xiang Zhang, Lina Yao, Dalin Zhang, Xianzhi Wang, Quan Z. Sheng, and Tao Gu. Multi-person
brain activity recognition via comprehensive EEG signal analysis. In *International Conference*
on Mobile and Ubiquitous Systems: Computing, Networking and Services, 2017.

-
- 810 Linfeng Zheng, Peilin Chen, and Shiqi Wang. EidetiCom: A cross-modal brain-computer semantic
811 communication paradigm for decoding visual perception. *arXiv*, 2407.14936, 2024a.
- 812
- 813 Xiao Zheng and Wanzhong Chen. An attention-based bi-LSTM method for visual object classifica-
814 tion via EEG. *Biomedical Signal Processing and Control*, 63, 2021.
- 815 Xiao Zheng, Wanzhong Chen, Mingyang Li, Tao Zhang, Yang You, and Yun Jiang. Decoding human
816 brain activity with deep learning. *Biomedical Signal Processing and Control*, 56, 2020a.
- 817
- 818 Xiao Zheng, Wanzhong Chen, Yang You, Yun Jiang, Mingyang Li, and Tao Zhang. Ensemble deep
819 learning for automated visual classification using EEG signals. *Pattern Recognition*, 102, 2020b.
- 820 Xu Zheng, Ling Wang, Kanghao Chen, Yuanhuiyi Lyu, Jiazhou Zhou, and Lin Wang. EIT-1M: One
821 million EEG-image-text pairs for human visual-textual recognition and more. *arXiv*, 2407.01884,
822 2024b.
- 823
- 824 Saisai Zhong, Yadong Liu, Zongtan Zhou, and Dewen Hu. ELSTM-based visual decoding from sin-
825 gular [sic]-trial EEG recording. In *International Conference on Software Engineering and Service
826 Science*, pp. 1139–1142, 2018.
- 827 Shuqi Zhu, Ziyi Ye, Qingyao Ai, and Yiqun Liu. EEG-ImageNet: An electroencephalogram dataset
828 and benchmarks with image visual stimuli of multi-granularity labels. *arXiv*, 2406.07151, 2024.
- 829

830 A APPENDIX

831 B METHOD

832

833 We report all results for three sets of EEG data: the data reported by Spampinato et al. (2017) as
834 used by Palazzo et al. (2018; 2020a;b; 2021) (‘Spampinato et al. (2017)’), the data reported by Li
835 et al. (2021) for subject 6 first image block run (‘Li et al. (2021) block’), and the data reported by
836 Li et al. (2021) for subject 6 image rapid-event run (‘Li et al. (2021) randomized’). We report all
837 results for four off-the-shelf image classifiers taken as image encoders: Inception v3 (Szegedy et al.,
838 2016), ResNet-101 (He et al., 2016), DenseNet-161 (Huang et al., 2017), and AlexNet (Krizhevsky
839 et al., 2012). We report all results both before joint training (‘before’), where the EEG encoder is
840 randomly initialized, after joint training (‘after joint’), and after separate training (‘after separate’).
841 For ‘before’ and ‘after separate’ the image encoder is initialized with off-the-shelf weights trained on
842 the ILSVRC 2012 training set (Russakovsky et al., 2014). For ‘joint’, we report results both where
843 the image encoder is initialized with off-the-shelf weights trained on the ILSVRC 2012 training set
844 (Russakovsky et al., 2014) (‘pretraining’) and where the image encoder is randomly initialized (‘no
845 pretraining’). The EEG and image encoders both output a 1000 element encoding. During joint
846 training, a triplet loss is used to drive the EEG and image encodings to have a high dot product
847 when the EEG signal is associated with the stimulus image that elicited the associated EEG signal
848 and a low dot product when the EEG signal is associated with a different image, irrespective of
849 class. We do not postpend a classifier (ReLU, 1000→40 FC, softmax) to the encoders, either during
850 training or after training. All analyses are done directly on the unmodified encodings produced by
851 the encoders. For both the EEG encodings (‘EEG’) and the image encodings (‘image’), we analyze
852 the unmodified encodings (‘original’) and encodings reconstructed by principal component analysis
853 (‘PCA’). For PCA, we analyze two reconstructions: one with only the top 40 principal components
854 (‘top 40’) and one with only the bottom 960 principal components (‘bottom 960’). We classify
855 original and reconstructed EEG and image encodings, before and after joint and separate training, in
856 two ways. In the first, we simply select the label of the maximal element among all 1000 elements
857 (‘1000 classes’). In the second, we select the label of the maximal element among only the 40
858 classes used as stimuli (‘40 classes’). In both cases, correct labels are considered positives. In the
859 former, negatives can result from incorrect labels, both within and outside the subset of 40 ILSVRC
860 classes used as stimuli. In the latter, negatives can result only from incorrect labels within the subset
861 of 40 ILSVRC classes used as stimuli.

862

863

864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917

Table 5: The top 40 principal components explain a large portion of the image-encoding variance ($\geq 63.9\%$) before training (column v). This together with Table 2 implies that the pretrained image encoders produce encodings that contain (primarily) class information on all data. This supports claim A and calls claim I into question.

		EEG				image				
		before	after	after	after	before	after	after	after	
			joint	joint	separate		joint	joint	separate	
			pretraining	no pretraining			pretraining	no pretraining		
		i	ii	iii	iv	v	vi	vii	viii	
Spampinato et al. (2017)	Inception v3	i	44.7	37.2	15.3	99.5	64.1	94.9	98.9	100.0
	ResNet-101	ii	46.5	36.2	17.7	99.5	82.0	98.9	98.2	96.8
	DenseNet-161	iii	48.2	25.6	22.7	100.0	75.4	97.4	90.2	100.0
	AlexNet	iv	46.3	21.7	18.5	99.5	82.3	nan	100.0	99.7
Li et al. (2021) block	Inception v3	v	83.3	75.9	69.4	42.3	63.9	57.4	98.3	99.9
	ResNet-101	vi	85.4	79.8	72.8	41.9	81.9	99.5	99.0	99.7
	DenseNet-161	vii	84.0	79.3	76.3	42.9	75.3	69.4	94.3	99.5
	AlexNet	viii	84.4	78.9	73.2	40.0	82.2	78.9	99.9	98.9
Li et al. (2021) randomized	Inception v3	ix	62.1	40.5	38.3	19.3	63.9	60.2	98.1	99.9
	ResNet-101	x	61.9	52.2	43.6	19.7	81.9	99.1	98.8	99.8
	DenseNet-161	xi	63.5	48.5	46.5	19.6	75.3	70.8	93.9	99.7
	AlexNet	xii	63.6	49.5	41.3	19.2	82.2	86.9	99.9	98.8

Table 6: After joint training with pretraining on the ‘Spampinato et al. (2017)’ data, the top 40 principal components explain almost all ($\geq 94.9\%$) of the variance in the image encodings (column vi, for rows i–iv). This implies that the image encoders jointly trained on the ‘Spampinato et al. (2017)’ data produce encodings with little more than class information.

		EEG				image				
		before	after	after	after	before	after	after	after	
			joint	joint	separate		joint	joint	separate	
			pretraining	no pretraining			pretraining	no pretraining		
		i	ii	iii	iv	v	vi	vii	viii	
Spampinato et al. (2017)	Inception v3	i	44.7	37.2	15.3	99.5	64.1	94.9	98.9	100.0
	ResNet-101	ii	46.5	36.2	17.7	99.5	82.0	98.9	98.2	96.8
	DenseNet-161	iii	48.2	25.6	22.7	100.0	75.4	97.4	90.2	100.0
	AlexNet	iv	46.3	21.7	18.5	99.5	82.3	nan	100.0	99.7
Li et al. (2021) block	Inception v3	v	83.3	75.9	69.4	42.3	63.9	57.4	98.3	99.9
	ResNet-101	vi	85.4	79.8	72.8	41.9	81.9	99.5	99.0	99.7
	DenseNet-161	vii	84.0	79.3	76.3	42.9	75.3	69.4	94.3	99.5
	AlexNet	viii	84.4	78.9	73.2	40.0	82.2	78.9	99.9	98.9
Li et al. (2021) randomized	Inception v3	ix	62.1	40.5	38.3	19.3	63.9	60.2	98.1	99.9
	ResNet-101	x	61.9	52.2	43.6	19.7	81.9	99.1	98.8	99.8
	DenseNet-161	xi	63.5	48.5	46.5	19.6	75.3	70.8	93.9	99.7
	AlexNet	xii	63.6	49.5	41.3	19.2	82.2	86.9	99.9	98.8

Table 7: After separate training, the top 40 principal components explain almost all of the variance ($\geq 96.8\%$) in the image encodings (column viii). This implies that the separately trained image encoders produce encodings with little more than class information on all data.

		EEG				image				
		before	after	after	after	before	after	after	after	
			joint	joint	separate		joint	joint	separate	
			pretraining	no pretraining			pretraining	no pretraining		
		i	ii	iii	iv	v	vi	vii	viii	
Spampinato et al. (2017)	Inception v3	i	44.7	37.2	15.3	99.5	64.1	94.9	98.9	100.0
	ResNet-101	ii	46.5	36.2	17.7	99.5	82.0	98.9	98.2	96.8
	DenseNet-161	iii	48.2	25.6	22.7	100.0	75.4	97.4	90.2	100.0
	AlexNet	iv	46.3	21.7	18.5	99.5	82.3	nan	100.0	99.7
Li et al. (2021) block	Inception v3	v	83.3	75.9	69.4	42.3	63.9	57.4	98.3	99.9
	ResNet-101	vi	85.4	79.8	72.8	41.9	81.9	99.5	99.0	99.7
	DenseNet-161	vii	84.0	79.3	76.3	42.9	75.3	69.4	94.3	99.5
	AlexNet	viii	84.4	78.9	73.2	40.0	82.2	78.9	99.9	98.9
Li et al. (2021) randomized	Inception v3	ix	62.1	40.5	38.3	19.3	63.9	60.2	98.1	99.9
	ResNet-101	x	61.9	52.2	43.6	19.7	81.9	99.1	98.8	99.8
	DenseNet-161	xi	63.5	48.5	46.5	19.6	75.3	70.8	93.9	99.7
	AlexNet	xii	63.6	49.5	41.3	19.2	82.2	86.9	99.9	98.8

918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971

Table 8: After separate training, the top 40 principal components explain almost all of the variance ($\geq 96.8\%$) in both the EEG and image encodings on the ‘Spampinato et al. (2017)’ data (columns iv and viii, for rows i–iv). This implies that the separately trained EEG and image encoders produce encodings with little more than class information on the ‘Spampinato et al. (2017)’ data.

			before	EEG		after	before	image		after
				after	joint	separate		after	joint	separate
				pretraining	no pretraining			pretraining	no pretraining	
			i	ii	iii	iv	v	vi	vii	viii
Spampinato et al. (2017)	Inception v3	i	44.7	37.2	15.3	99.5	64.1	94.9	98.9	100.0
	ResNet-101	ii	46.5	36.2	17.7	99.5	82.0	98.9	98.2	96.8
	DenseNet-161	iii	48.2	25.6	22.7	100.0	75.4	97.4	90.2	100.0
	AlexNet	iv	46.3	21.7	18.5	99.5	82.3	nan	100.0	99.7
Li et al. (2021) block	Inception v3	v	83.3	75.9	69.4	42.3	63.9	57.4	98.3	99.9
	ResNet-101	vi	85.4	79.8	72.8	41.9	81.9	99.5	99.0	99.7
	DenseNet-161	vii	84.0	79.3	76.3	42.9	75.3	69.4	94.3	99.5
	AlexNet	viii	84.4	78.9	73.2	40.0	82.2	78.9	99.9	98.9
Li et al. (2021) randomized	Inception v3	ix	62.1	40.5	38.3	19.3	63.9	60.2	98.1	99.9
	ResNet-101	x	61.9	52.2	43.6	19.7	81.9	99.1	98.8	99.8
	DenseNet-161	xi	63.5	48.5	46.5	19.6	75.3	70.8	93.9	99.7
	AlexNet	xii	63.6	49.5	41.3	19.2	82.2	86.9	99.9	98.8

972
 973
 974
 975
 976
 977
 978
 979
 980
 981
 982
 983
 984
 985
 986
 987
 988
 989
 990
 991
 992
 993
 994
 995
 996
 997
 998
 999
 1000
 1001
 1002
 1003
 1004
 1005
 1006
 1007
 1008
 1009
 1010
 1011
 1012
 1013
 1014
 1015
 1016
 1017
 1018
 1019
 1020
 1021
 1022
 1023
 1024
 1025

Table 9: Image classification accuracy is near perfect with all components with 40 classes ($\geq 95.5\%$; **column i**, for rows v–viii, xiii–xvi, and xxi–xxiv) before training. This holds both for the ‘Spampinato et al. (2017)’ data, the ‘Li et al. (2021) block’ data, and the ‘Li et al. (2021) randomized’ data. This, together with Tables 10 and 11, demonstrates that they are giving class information implicitly to their training by use of image encoders pretrained on ImageNet, supports claim A, and calls claim I into question.

		original				bottom 960				40-classes				1000-classes				bottom 960				after separate												
		before	pretraining	no pretraining	joint	before	pretraining	no pretraining	joint	before	pretraining	no pretraining	joint	before	pretraining	no pretraining	joint	before	pretraining	no pretraining	joint	before	pretraining	no pretraining	joint	before	pretraining	no pretraining	joint					
Spampinato et al. (2017)	image	i	2.6	3.0	2.5	2.2	2.3	2.3	3.1	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3			
		ii	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5			
		iii	2.5	3.1	2.4	57.4	2.4	2.4	2.4	2.5	2.4	2.5	2.4	2.5	2.4	2.5	2.4	2.5	2.4	2.5	2.4	2.5	2.4	2.5	2.4	2.5	2.4	2.5	2.4	2.5	2.4	2.5		
		iv	2.3	2.5	2.4	57.3	2.4	2.6	2.6	2.5	2.6	2.5	2.6	2.5	2.6	2.5	2.6	2.5	2.6	2.5	2.6	2.5	2.6	2.5	2.6	2.5	2.6	2.5	2.6	2.5	2.6	2.5		
	BEG	v	99.7	43.2	2.7	34	99.4	13.5	2.9	2.2	45.5	29.4	2.7	2.5	92.8	16.5	0.1	0.0	98.9	4.0	0.0	100	12.8	9.3	0.0	60.0	0.0	0.0	0.0	0.0	0.0	30.0	0.0	
		vi	99.4	24.2	2.4	23.5	97.5	4.3	2.9	7.5	4.0	5.8	2.1	39.4	89.4	2.6	0.0	92.3	5.6	0.0	100	12.8	9.3	0.0	30.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
		vii	99.4	24.2	2.4	23.5	97.5	4.3	2.9	7.5	4.0	5.8	2.1	39.4	89.4	2.6	0.0	92.3	5.6	0.0	100	12.8	9.3	0.0	30.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		viii	95.8	2.6	2.7	18.7	85.8	2.6	2.7	19.1	34.1	42.7	2.4	2.4	78.4	0.0	0.3	18.7	72.5	0.0	0.3	19.1	12.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Li et al. (2021) block	BEG	ix	2.4	6.8	1.9	77.0	1.8	7.1	2.0	69.8	2.6	3.6	2.5	32.1	0.1	0.4	0.0	77.0	0.0	0.2	0.1	68.8	0.0	0.2	0.1	28.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
		x	1.6	3.5	2.2	76.6	1.1	4.2	2.5	70.5	2.4	2.8	2.2	28.1	0.0	0.1	0.1	76.6	0.0	0.1	0.1	70.5	0.0	0.0	0.0	0.1	24.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		xi	2.4	4.1	2.0	76.0	0.2	3.8	2.0	70.2	2.5	3.4	2.5	27.9	0.1	0.1	0.1	77.0	0.0	0.3	0.2	70.0	0.0	0.1	0.1	24.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		xii	2.4	4.1	2.0	76.0	0.2	3.8	2.0	70.2	2.5	3.4	2.5	27.9	0.1	0.1	0.1	77.0	0.0	0.3	0.2	70.0	0.0	0.1	0.1	24.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	image	xiii	99.3	96.4	3.0	93	99.3	94.7	3.1	2.9	46.0	46.8	2.4	2.5	93.0	75.1	0.1	0.5	98.8	87.0	0.0	0.0	13.2	11.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		xiv	99.3	75.0	2.4	11.0	98.5	68.5	2.7	3.5	34.1	14.7	2.5	3.7	90.5	56.4	0.0	2.0	96.2	52.4	0.0	0.1	9.8	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		xv	99.1	95.8	2.5	5.7	97.7	86.8	2.9	3.8	43.9	50.9	2.1	2.6	89.5	71.1	0.1	0.2	96.4	69.2	0.4	0.2	15.0	17.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		xvi	95.5	65.9	2.5	1.5	85.7	52.9	2.4	2.2	34.6	19.6	2.7	1.5	71.9	25.4	0.0	0.1	85.7	44.8	0.0	0.0	2.5	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Li et al. (2021) randomized	BEG	xvii	2.6	2.4	2.8	2.1	2.8	2.7	2.9	2.3	2.6	2.7	3.0	1.8	0.0	0.1	1.9	0.0	0.1	1.9	0.0	0.1	0.2	2.3	0.1	0.2	0.1	1.6	0.0	0.0	0.0	0.0	0.0	
		xviii	2.1	2.2	2.3	1.5	2.5	2.3	2.5	2.0	3.1	2.0	1.9	1.9	0.1	0.1	1.5	0.0	0.0	0.1	2.0	0.1	0.1	0.1	0.1	1.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
		xix	2.5	2.8	3.0	2.2	2.2	2.4	2.2	1.6	2.6	2.8	3.1	2.4	0.2	0.1	2.2	0.3	0.0	0.1	1.6	0.1	0.1	0.1	0.1	1.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
		xx	99.3	69.8	2.9	11.2	98.5	53.1	2.9	4.0	40.0	42.0	2.1	2.2	90.5	42.0	0.0	2.0	96.2	28.5	0.0	0.1	9.8	3.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Li et al. (2021) randomized	BEG	xxi	99.1	65.6	2.1	5.1	97.7	36.0	2.5	3.0	43.9	36.1	2.3	3.0	89.5	30.5	0.1	0.0	96.4	11.7	0.0	0.0	15.0	13.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
		xxii	99.1	65.6	2.1	5.1	97.7	36.0	2.5	3.0	43.9	36.1	2.3	3.0	89.5	30.5	0.1	0.0	96.4	11.7	0.0	0.0	15.0	13.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
		xxiii	95.5	65.6	2.1	5.1	97.7	36.0	2.5	3.0	43.9	36.1	2.3	3.0	89.5	30.5	0.1	0.0	96.4	11.7	0.0	0.0	15.0	13.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
		xxiv	95.5	65.6	2.1	5.1	97.7	36.0	2.5	3.0	43.9	36.1	2.3	3.0	89.5	30.5	0.1	0.0	96.4	11.7	0.0	0.0	15.0	13.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079

Table 10: Image classification accuracy is very high with all components with 1000 classes ($\geq 77.9\%$; column xiii, for rows v–viii, xiii–xvi, and xxi–xxiv) before training. This holds both for the ‘Spampinato et al. (2017)’ data, the ‘Li et al. (2021) block’ data, and the ‘Li et al. (2021) randomized’ data. This, together with Tables 9 and 11, demonstrates that they are giving class information implicitly to their training by use of image encoders pretrained on ImageNet, supports claim A, and calls claim 1 into question.

		40-classes top-40				bottom 960				original joint				1000-classes top-40				bottom 960				
		before	after separate	before	after separate	before	after separate	before	after separate	before	after separate	before	after separate	before	after separate	before	after separate	before	after separate			
Spampinato et al. (2017)	inception v3	i	2.6	2.5	2.2	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3		
	ResNet-101	ii	2.5	2.4	2.2	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3		
	DenseNet-161	iii	2.5	2.4	2.2	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3		
	AlexNet	iv	2.3	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4		
	inception v3	v	99.7	43.2	2.7	3.4	99.4	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7		
	ResNet-101	vi	99.4	2.2	2.4	2.3	97.3	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2		
	DenseNet-161	vii	99.4	2.2	2.4	2.3	97.3	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2		
	AlexNet	viii	95.8	2.6	2.7	18.7	85.8	2.6	2.7	19.1	34.1	2.6	2.4	2.4	78.4	0.0	0.3	18.7	72.5	0.0	0.0	
Li et al. (2021) block	inception v3	ix	2.4	6.8	1.9	77.0	1.8	1.8	2.0	69.8	2.6	3.6	2.5	32.1	0.1	0.1	77.0	0.0	0.2	0.1		
	ResNet-101	x	1.6	3.5	2.2	76.6	1.1	1.1	2.5	70.5	2.4	2.8	2.2	28.1	0.0	0.1	76.6	0.0	0.1	0.1		
	DenseNet-161	xi	2.0	4.1	2.0	76.0	1.2	1.2	2.5	70.2	2.5	3.4	2.5	27.9	0.1	0.1	76.0	0.0	0.1	0.1		
	AlexNet	xii	2.4	4.1	2.0	76.0	1.2	1.2	2.5	70.2	2.5	3.4	2.5	27.9	0.1	0.1	76.0	0.0	0.1	0.1		
	inception v3	xiii	97.8	96.4	3.0	93.3	99.3	94.7	3.1	2.9	46.0	46.8	2.4	2.3	93.0	75.1	0.1	98.8	87.0	0.0	0.0	
	ResNet-101	xiv	99.3	75.0	2.4	11.0	98.5	68.5	2.7	3.5	34.1	14.7	2.5	3.7	90.5	56.4	0.0	2.0	96.2	52.4	0.0	0.0
	DenseNet-161	xv	99.1	95.8	2.5	5.7	97.7	86.8	2.9	3.8	43.9	50.9	2.1	2.6	89.5	71.1	0.1	0.2	96.4	69.2	0.4	0.2
	AlexNet	xvi	92.5	62.9	2.5	1.5	85.7	52.9	2.4	2.2	34.6	19.6	2.7	1.3	71.9	20.4	0.1	0.1	71.9	14.4	0.1	0.1
Li et al. (2021) randomized	inception v3	xvii	2.6	2.4	2.8	2.1	2.8	2.7	2.9	2.3	2.6	2.7	3.0	1.8	0.0	0.1	1.9	0.0	0.1	0.2		
	ResNet-101	xviii	2.6	2.4	2.8	2.1	2.8	2.7	2.9	2.3	2.6	2.7	3.0	1.8	0.0	0.1	1.9	0.0	0.1	0.2		
	DenseNet-161	xix	2.1	2.2	2.3	1.5	2.5	2.0	3.1	2.0	3.1	2.0	1.9	1.9	0.1	1.5	0.0	0.1	1.5	0.0		
	AlexNet	xx	2.5	2.8	3.0	2.2	2.2	2.2	2.4	2.9	1.6	2.6	2.8	3.1	2.4	0.2	0.1	2.2	0.3	0.0	0.1	
	inception v3	xxi	99.3	69.8	2.9	11.2	98.5	53.1	2.9	4.0	34.1	9.4	3.1	2.2	90.5	42.0	0.0	2.0	96.2	28.5	0.0	0.0
	ResNet-101	xxii	99.3	69.8	2.9	11.2	98.5	53.1	2.9	4.0	34.1	9.4	3.1	2.2	90.5	42.0	0.0	2.0	96.2	28.5	0.0	0.0
	DenseNet-161	xxiii	65.6	2.1	5.1	97.7	36.0	2.5	3.0	43.9	36.1	2.3	3.0	89.5	30.5	0.1	0.0	15.0	13.6	0.0	0.0	
	AlexNet	xxiv	95.5	2.4	3.5	85.7	34.6	2.5	2.5	2.5	34.6	2.5	2.5	2.5	77.9	0.0	0.0	12.9	0.0	0.0	0.0	

1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133

Table 11: Image classification accuracy is very high with just the top 40 principal components both with 40 classes ($\geq 85.7\%$; column v, for rows v–viii, xiii–xvi, and xxi–xxiv) and 1000 classes ($\geq 72.2\%$; column xvii, for rows v–viii, xiii–xvi, and xxi–xxiv). This holds both for the ‘Spampinato et al. (2017)’ data, the ‘Li et al. (2021) block’ data, and the ‘Li et al. (2021) randomized’ data. This, together with Tables 9 and 10, demonstrates that they are giving class information implicitly to their training by use of image encoders pretrained on ImageNet, supports claim A, and calls claim I into question.

	40-classes				1000-classes				
	before	after separate	after joint	no pretraining	before	after separate	after joint	no pretraining	
Spampinato et al. (2017)	i	2.6	2.5	2.6	2.6	2.3	2.3	2.3	2.3
	ii	3.0	3.4	3.1	3.1	3.3	3.3	3.1	3.3
	iii	2.5	2.4	2.4	2.4	2.3	2.3	2.3	2.3
	iv	2.3	2.4	2.4	2.4	2.5	2.5	2.5	2.5
	v	99.4	99.4	99.4	99.4	98.9	98.9	98.9	98.9
	vi	99.4	99.4	99.4	99.4	98.9	98.9	98.9	98.9
	vii	99.4	99.4	99.4	99.4	98.9	98.9	98.9	98.9
	viii	95.8	95.8	95.8	95.8	95.8	95.8	95.8	95.8
Li et al. (2021) block	ix	2.4	1.9	1.8	1.8	2.1	2.1	2.1	2.1
	x	6.8	2.2	2.2	2.2	3.6	3.6	3.6	3.6
	xi	3.5	2.2	2.2	2.2	2.8	2.8	2.8	2.8
	xii	4.1	2.0	2.0	2.0	3.4	3.4	3.4	3.4
	xiii	96.4	96.4	96.4	96.4	95.8	95.8	95.8	95.8
	xiv	99.3	99.3	99.3	99.3	98.5	98.5	98.5	98.5
	xv	99.1	99.1	99.1	99.1	98.5	98.5	98.5	98.5
	xvi	95.5	95.5	95.5	95.5	95.5	95.5	95.5	95.5
Li et al. (2021) randomized	xvii	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
	xviii	2.6	2.8	2.1	2.8	2.7	2.7	2.7	2.7
	xix	2.1	2.2	2.3	2.3	2.0	2.0	2.0	2.0
	xx	2.8	3.0	2.2	2.2	2.6	2.6	2.6	2.6
	xxi	98.5	98.5	98.5	98.5	98.5	98.5	98.5	98.5
	xxii	99.3	99.3	99.3	99.3	98.5	98.5	98.5	98.5
	xxiii	65.6	65.6	65.6	65.6	65.6	65.6	65.6	65.6
	xxiv	95.5	95.5	95.5	95.5	95.5	95.5	95.5	95.5

1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187

Table 12: Image classification accuracy is much lower than perfect ($\leq 46.0\%$) before training when discarding the top 40 principal components (columns ix and xxi, for rows v–viii, xiii–xvi, and xxi–xxiv). This holds both for 40 classes and 1000 classes and both for the ‘Spampinato et al. (2017)’ data, the ‘Li et al. (2021) block’ data, and the ‘Li et al. (2021) randomized’ data. Thus there is much less class information outside the top 40 principal components. This demonstrates that they are giving little more than class information implicitly to their training by use of image encoders pretrained on ImageNet and further supports claim A and calls claim I into question.

		40-classes sp-40				bottom 960				1000-classes sp-1000				original data				
		after separate		joint		after separate		joint		after separate		joint		after separate		joint		
		no pretraining		pretraining		no pretraining		pretraining		no pretraining		pretraining		no pretraining		pretraining		
		iii	iv	v	vi	vii	viii	ix	x	xi	xii	xiii	xiv	xv	xvi	xvii	xviii	
Spampinato et al. (2017)	inception v3	2.5	2.5	2.4	2.6	2.3	2.3	2.5	2.5	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	
	ResNet-101	3.0	5.7	2.4	2.6	2.3	2.4	2.4	2.4	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	
	DenseNet-161	2.5	2.4	2.4	2.6	2.3	2.4	2.5	3.2	2.4	2.5	2.5	2.5	2.5	2.5	2.5	2.5	
	AlexNet	2.3	2.4	2.4	2.6	2.6	2.6	2.4	2.5	2.6	2.5	2.5	2.5	2.5	2.5	2.5	2.5	
	image	99.7	3.4	99.4	13.3	2.9	2.2	45.3	29.4	2.7	2.3	92.8	16.5	0.1	0.0	0.0	0.0	
	inception v3	43.2	3.4	99.4	13.3	2.9	2.2	45.3	29.4	2.7	2.3	92.8	16.5	0.1	0.0	0.0	0.0	
	ResNet-101	48.2	2.7	3.4	13.3	2.9	2.2	45.3	29.4	2.7	2.3	92.8	16.5	0.1	0.0	0.0	0.0	
	DenseNet-161	7.2	2.4	23.3	97.9	4.4	3.5	43.7	3.8	2.6	2.4	89.4	0.1	0.0	0.0	0.0	0.0	
	AlexNet	95.8	2.6	2.7	18.7	85.8	2.6	19.1	34.1	2.6	2.4	78.4	0.0	0.3	18.7	72.5	0.0	0.0
	inception v3	2.4	6.8	1.9	77.0	1.8	7.1	2.0	69.8	2.6	3.6	2.5	32.1	0.1	0.0	77.0	0.0	0.2
	ResNet-101	1.6	3.5	2.2	76.6	1.1	4.2	2.5	70.5	2.4	2.8	2.2	28.1	0.0	0.1	76.6	0.0	0.1
	DenseNet-161	99.1	65.8	2.5	5.7	97.7	56.8	2.9	3.8	45.9	2.1	2.6	89.5	71.1	0.1	0.2	96.4	69.2
AlexNet	2.9	4.1	2.9	76.5	2.5	2.2	70.0	3.4	2.5	27.8	0.0	0.1	76.4	0.0	0.3	70.0	0.0	
image	99.8	96.4	3.0	93.3	94.7	3.1	2.9	46.0	46.8	2.4	2.5	93.0	78.1	0.1	0.5	98.8	87.0	
ResNet-101	99.3	75.0	2.4	11.0	98.5	68.5	2.7	3.5	34.1	14.7	2.5	3.7	90.5	56.4	0.0	2.0	96.2	52.4
DenseNet-161	99.1	95.8	2.5	5.7	97.7	56.8	2.9	3.8	45.9	2.1	2.6	89.5	71.1	0.1	0.2	96.4	69.2	
AlexNet	2.6	2.5	2.5	76.5	2.5	2.2	70.0	3.4	2.5	27.8	0.0	0.1	76.4	0.0	0.3	70.0	0.0	
inception v3	2.6	2.5	2.5	1.6	2.6	2.6	2.5	3.0	3.0	2.7	1.6	0.1	0.1	1.5	0.1	0.1	2.5	
ResNet-101	2.6	2.4	2.8	2.1	2.8	2.7	2.9	2.3	2.6	2.7	3.0	1.8	0.0	0.1	1.9	0.0	0.1	
DenseNet-161	2.1	2.2	2.3	1.5	2.5	2.3	2.5	2.0	3.1	1.9	1.9	0.1	0.1	1.5	0.0	0.0	0.1	
AlexNet	3.0	3.0	3.0	2.2	3.0	2.4	2.9	2.6	2.6	3.1	2.4	0.2	0.1	0.2	2.2	0.3	0.1	
image	99.5	92.8	2.9	11.2	98.5	53.1	2.6	4.0	34.1	9.4	3.1	2.2	90.5	42.0	0.0	2.0	96.2	26.5
ResNet-101	99.3	69.8	2.9	11.2	98.5	53.1	2.6	4.0	34.1	9.4	3.1	2.2	90.5	42.0	0.0	2.0	96.2	26.5
DenseNet-161	99.1	65.6	2.1	5.1	97.7	36.0	2.5	3.0	43.9	36.1	2.3	3.0	89.5	30.5	0.1	0.0	96.4	11.7
AlexNet	95.5	3.6	2.4	85.7	3.6	2.5	2.5	2.5	3.5	34.6	2.5	2.5	77.9	0.0	0.0	2.6	72.2	

1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241

Table 13: With only five exceptions, all in the bottom 960 principal components, image classification accuracy decreases after joint training with pretraining (compare columns ii, vi, x, xiv, xviii, and xxii to i, v, ix, xiii, xvii, and xxi, respectively, for rows v–viii, xiii–xvi, and xxi–xxiv). This holds both for 40 classes and 1000 classes, both for the ‘Spampinato et al. (2017)’ data, the ‘Li et al. (2021)’ block’ data, and the ‘Li et al. (2021)’ randomized’ data, and whether using all components or just the 40 principal components. This suggests that joint training is hurting, not helping, and supports claim C(2).

	40 classes												1000 classes											
	before		after separate		original joint		before separate		after separate		original joint		before separate		after separate		bottom 960 joint		before separate		after separate			
	pretraining	no pretraining	pretraining	no pretraining	pretraining	no pretraining	pretraining	no pretraining	pretraining	no pretraining	pretraining	no pretraining	pretraining	no pretraining	pretraining	no pretraining	pretraining	no pretraining	pretraining	no pretraining	pretraining	no pretraining		
Spampinato et al. (2017)	i	2.6	3.0	2.5	2.8	2.5	2.8	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	
	ii	2.5	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	
	iii	2.5	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	
	iv	2.3	2.4	2.4	2.4	2.4	2.4	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	
	v	99.7	43.2	2.7	3.4	99.4	2.2	45.5	29.4	2.7	2.5	92.8	0.1	0.0	100	12.8	9.3	0.0	0.0	0.0	0.0	0.0	0.0	
	vi	99.7	43.2	2.7	3.4	99.4	2.2	45.5	29.4	2.7	2.5	92.8	0.1	0.0	100	12.8	9.3	0.0	0.0	0.0	0.0	0.0	0.0	
	vii	99.7	43.2	2.7	3.4	99.4	2.2	45.5	29.4	2.7	2.5	92.8	0.1	0.0	100	12.8	9.3	0.0	0.0	0.0	0.0	0.0	0.0	
	viii	95.8	2.6	2.7	18.7	85.8	2.6	34.1	2.6	2.4	78.4	0.0	0.3	18.7	72.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Li et al. (2021) block	ix	2.4	6.8	1.9	77.0	1.8	2.0	69.8	2.6	3.6	32.1	0.0	0.4	0.0	77.0	0.0	0.2	0.1	68.8	0.0	0.2	0.1	28.0	
	x	1.6	3.5	2.2	76.6	1.1	4.2	2.5	70.5	2.4	2.8	2.2	28.1	0.0	0.1	76.6	0.0	0.1	70.5	0.0	0.0	0.1	24.0	
	xi	2.0	4.1	2.0	76.0	1.2	3.8	2.9	70.2	2.5	2.7	2.0	28.1	0.0	0.1	76.0	0.0	0.1	70.2	0.0	0.0	0.1	24.0	
	xi	2.4	4.1	2.0	76.0	1.2	3.8	2.9	70.2	2.5	2.7	2.0	28.1	0.0	0.1	76.0	0.0	0.1	70.2	0.0	0.0	0.1	24.2	
	xiii	99.3	96.4	3.0	93.3	99.3	3.1	2.9	46.0	46.8	2.4	2.5	93.0	75.1	0.1	0.5	98.8	87.0	0.0	100	13.2	0.0	0.0	
	xiv	99.3	96.4	3.0	93.3	99.3	3.1	2.9	46.0	46.8	2.4	2.5	93.0	75.1	0.1	0.5	98.8	87.0	0.0	100	13.2	0.0	0.0	
	xv	99.1	95.8	2.5	57	97.7	2.5	3.8	43.9	50.9	2.1	2.6	89.5	71.1	0.1	0.2	96.4	62.2	0.4	0.2	15.0	17.9	0.0	
	xvi	99.5	95.9	2.5	57	97.7	2.5	3.8	43.9	50.9	2.1	2.6	89.5	71.1	0.1	0.2	96.4	62.2	0.4	0.2	15.0	17.9	0.0	
	xvii	2.6	2.4	2.8	2.1	2.8	2.7	2.9	2.3	2.6	2.7	3.0	1.8	0.0	0.1	1.9	0.0	0.1	2.0	0.1	0.1	0.1	1.6	
	xviii	2.1	2.2	2.3	1.5	2.5	2.0	3.1	2.0	3.1	1.9	1.9	0.1	0.1	1.5	0.0	0.0	0.1	2.0	0.1	0.1	0.1	1.7	
	xix	2.5	2.8	3.0	2.2	2.2	2.2	2.2	1.6	2.6	2.8	3.1	2.4	0.2	0.1	2.2	0.3	0.0	0.1	1.6	0.1	0.1	1.9	
	xx	99.3	69.8	2.9	11.2	98.5	51.7	2.6	40	34.1	3.1	2.2	90.5	42.0	0.0	2.0	96.2	28.5	0.0	0.1	9.8	3.5	0.0	
	xxi	99.3	69.8	2.9	11.2	98.5	51.7	2.6	40	34.1	3.1	2.2	90.5	42.0	0.0	2.0	96.2	28.5	0.0	0.1	9.8	3.5	0.0	
	xxii	99.1	65.6	2.1	5.1	97.7	36.0	2.5	30	43.9	36.1	2.3	3.0	89.5	30.5	0.1	0.0	96.4	11.7	0.0	0.0	13.6	0.0	
	xxiii	99.1	65.6	2.1	5.1	97.7	36.0	2.5	30	43.9	36.1	2.3	3.0	89.5	30.5	0.1	0.0	96.4	11.7	0.0	0.0	13.6	0.0	
	xxiv	95.5	3.6	2.4	85.7	85.7	2.5	3.5	34.6	2.5	2.5	77.9	0.0	0.0	2.6	72.2	0.0	0.0	2.6	12.9	0.0	0.0	2.5	

1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295

Table 14: With the exception of a small number of cases that are marginally above chance, EEG classification accuracy is at chance except after separate training on confounded data either on all components or the top 40 components (columns iv, viii, xvi, and xx, for rows i-iv and ix-xii). This, together with Table 15, suggests that separate training is able to extract class from confounded data but not nonconfounded data and that joint training is not able to extract class from any data, and supports claims C(1) and D.

		40-classes top-40				bottom 960				original joint				1000-classes top-40				bottom 960						
		before	after separate	before	after separate	before	after separate	before	after separate	before	after separate	before	after separate	before	after separate	before	after separate	before	after separate					
Spampinato et al. (2017)	EEG	i	2.6	2.5	2.2	2.3	3.1	2.3	2.3	2.3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1				
		ii	3.0	2.5	2.2	2.3	3.1	2.3	2.3	2.3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1				
		iii	2.5	2.4	2.2	2.3	3.1	2.3	2.3	2.3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1			
		iv	2.3	2.4	2.2	2.3	3.1	2.3	2.3	2.3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1			
	image	v	99.7	43.2	2.7	3.4	99.4	13.5	2.9	2.2	45.5	29.4	2.7	2.5	92.8	16.5	0.1	0.0	100	12.8	0.0	60.0		
		vi	99.4	24.2	2.4	2.4	97.5	4.3	2.9	2.2	35.5	3.8	2.1	39.4	2.6	0.0	0.0	100	12.8	0.0	30.0			
		vii	99.4	24.2	2.4	2.4	97.5	4.3	2.9	2.2	35.5	3.8	2.1	39.4	2.6	0.0	0.0	100	12.8	0.0	30.0			
		viii	95.8	2.6	2.7	18.7	85.8	2.6	2.7	19.1	34.1	2.4	2.4	78.4	0.0	0.3	18.7	72.5	0.0	0.0	0.0	2.4		
Tu et al. (2021) ResNet	EEG	ix	2.4	6.8	1.9	77.0	1.8	7.1	2.0	69.8	2.6	32.1	0.1	0.4	0.0	77.0	0.0	0.2	0.1	69.8	0.0	0.2	0.1	
		x	1.6	3.5	2.2	76.6	1.1	4.2	2.5	70.5	2.4	28.1	0.0	0.1	0.1	76.6	0.0	0.1	0.1	70.5	0.0	0.0	0.1	
		xi	4.1	2.0	2.0	76.0	2.2	3.5	2.3	70.2	2.0	27.9	0.1	0.1	0.1	76.0	0.0	0.1	0.1	70.2	0.0	0.1	0.1	
		xii	2.4	4.1	2.0	76.0	2.2	3.5	2.3	70.2	2.0	27.9	0.1	0.1	0.1	76.0	0.0	0.1	0.1	70.2	0.0	0.1	0.1	
	image	xiii	97.8	96.4	3.0	93.3	99.3	94.7	3.1	2.9	46.0	46.8	2.4	2.5	93.0	75.1	0.1	0.5	98.8	87.0	0.0	0.0	0.0	
		xiv	99.3	75.0	2.4	11.0	98.5	68.5	2.7	3.5	34.1	14.7	2.5	3.7	90.5	56.4	0.0	2.0	96.2	52.4	0.0	0.1	9.8	
		xv	99.1	95.8	2.5	5.7	97.7	86.8	2.9	3.8	43.9	50.9	2.1	2.6	89.5	71.1	0.1	0.2	96.4	69.2	0.4	0.2	15.0	
		xvi	92.5	62.9	2.5	1.5	85.7	52.9	2.4	2.2	34.6	19.6	2.0	2.1	85.7	52.9	0.1	0.1	85.7	52.9	0.1	0.1	17.9	
Tu et al. (2021) randomized	EEG	xvii	2.6	2.4	2.8	2.1	2.8	2.7	2.9	2.3	2.6	2.7	3.0	1.8	0.0	0.1	1.9	0.0	0.1	0.0	2.3	0.1	0.2	0.1
		xviii	2.1	2.2	2.3	1.5	2.5	2.3	2.5	2.0	3.1	2.0	1.9	1.9	0.1	0.1	1.5	0.0	0.0	0.1	2.0	0.1	0.1	0.1
		xix	2.1	2.2	2.3	1.5	2.5	2.3	2.5	2.0	3.1	2.0	1.9	1.9	0.1	0.1	1.5	0.0	0.0	0.1	2.0	0.1	0.1	0.1
		xx	2.5	2.8	3.0	2.2	2.2	2.4	2.2	1.6	2.6	2.8	3.1	2.4	0.2	0.1	2.2	0.3	0.0	0.1	1.6	0.1	0.1	0.1
	image	xxi	99.3	69.8	2.9	11.2	98.5	53.1	2.9	40.0	34.1	30.0	2.1	2.2	90.5	42.0	0.0	2.0	96.2	28.5	0.0	0.1	9.8	
		xxii	99.1	65.6	2.1	5.1	97.7	36.0	2.5	3.0	43.9	36.1	2.3	3.0	89.5	30.5	0.1	0.0	96.4	11.7	0.0	0.0	13.6	
		xxiii	99.1	65.6	2.1	5.1	97.7	36.0	2.5	3.0	43.9	36.1	2.3	3.0	89.5	30.5	0.1	0.0	96.4	11.7	0.0	0.0	13.6	
		xxiv	95.5	3.6	2.4	3.5	85.7	3.6	2.5	2.5	34.6	2.5	2.5	77.9	0.0	0.0	2.6	72.2	0.0	0.0	2.6	12.9	0.0	0.0

1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349

Table 15: EEG classification accuracy is also above chance after separate training on the ‘Li et al. (2021) block’ data in the bottom 960 components (columns **xii** and **xxiv**, for rows **ix—xii**). This, together with Table 14, suggests that separate training is able to extract class from confounded data but not nonconfounded data and that joint training is not able to extract class from any data, and supports claims C(1) and D.

	before		original joint		top-40 joint		bottom 960 joint		1000 classes joint		before		bottom 960 joint		after		
	pretraining	no pretraining	pretraining	no pretraining	pretraining	no pretraining	pretraining	no pretraining	pretraining	no pretraining	pretraining	no pretraining	pretraining	no pretraining	separate	joint	
Spampinato et al. (2017)	i	2.6	2.8	3.0	2.2	2.6	2.3	3.1	2.5	2.3	0.1	3.1	0.1	3.1	0.1	3.1	2.3
	ii	2.5	2.9	3.1	2.4	2.6	2.3	3.2	2.4	2.3	0.1	3.1	0.1	3.1	0.1	3.1	2.3
	iii	2.3	2.4	2.5	2.4	2.6	2.3	3.2	2.4	2.3	0.1	3.1	0.1	3.1	0.1	3.1	2.5
	iv	2.3	2.4	2.5	2.4	2.6	2.3	3.2	2.4	2.3	0.1	3.1	0.1	3.1	0.1	3.1	2.5
	v	99.7	43.2	2.7	3.4	99.4	13.5	2.9	2.2	48.5	28.4	2.7	2.5	92.8	0.1	0.1	0.0
	vi	99.4	24.2	2.4	82.5	98.3	7.3	2.9	70.5	33.0	9.5	2.1	39.4	90.5	2.6	0.0	82.5
	vii	95.8	2.6	2.7	18.7	85.8	2.6	2.7	19.3	34.1	2.6	2.4	78.4	0.0	0.3	18.7	72.5
	viii	95.8	2.6	2.7	18.7	85.8	2.6	2.7	19.3	34.1	2.6	2.4	78.4	0.0	0.3	18.7	72.5
	ix	2.4	6.8	1.9	77.0	1.8	7.1	2.0	69.8	2.6	2.5	32.1	0.1	0.4	0.0	77.0	0.0
	x	1.6	3.5	2.2	76.6	1.1	4.2	2.5	70.5	2.4	2.8	28.1	0.0	0.1	0.1	76.6	0.0
	xi	1.4	4.1	2.0	77.0	1.2	3.8	1.9	72.2	2.5	3.5	27.7	0.1	0.1	0.0	77.0	0.0
	Li et al. (2021) block	xii	97.8	96.4	3.0	93.3	97.3	98.7	3.1	29.9	46.0	46.8	2.4	2.3	93.0	75.1	0.1
xiii		99.3	75.0	2.4	11.0	98.5	68.5	2.7	3.5	34.1	14.7	2.5	3.7	90.5	56.4	0.0	2.0
xiv		99.1	95.8	2.5	5.7	97.7	86.8	2.9	3.8	43.9	50.9	2.1	2.6	89.5	71.1	0.1	0.2
xv		95.5	60.9	2.5	3.5	85.7	50.9	2.4	3.2	34.6	15.6	2.3	77.9	25.4	0.0	2.6	
xvi		2.5	2.8	2.1	2.8	2.1	2.8	2.9	2.3	2.6	2.7	3.0	1.8	0.0	0.1	1.9	0.0
xvii		2.1	2.2	2.3	1.5	2.5	2.3	2.5	2.0	3.1	2.0	1.9	1.9	0.1	1.5	0.0	0.1
xviii		2.6	2.4	2.8	2.1	2.8	2.1	2.9	2.3	2.6	2.7	3.0	1.8	0.0	0.1	1.9	0.0
xix		2.5	2.8	3.0	2.2	2.2	2.4	2.9	1.6	2.6	2.8	3.1	2.4	0.2	2.2	0.3	0.0
xx		2.5	2.8	3.0	2.2	2.2	2.4	2.9	1.6	2.6	2.8	3.1	2.4	0.2	2.2	0.3	0.0
xxi		99.3	99.3	3.4	98.4	98.4	81.7	3.0	40.0	30.1	31.1	25.3	39.0	40.0	0.1	30.3	28.5
xxii		99.1	65.6	2.1	11.1	98.4	69.3	2.5	4.0	33.0	44.0	3.1	2.7	90.5	44.0	0.0	0.1
Lu et al. (2021) randomized		xxiii	99.1	65.6	2.1	51.1	97.7	36.0	2.5	3.0	43.9	36.1	2.3	3.0	89.5	30.5	0.1
	xxiv	95.5	2.4	3.5	85.7	3.6	2.5	2.5	34.6	2.5	2.5	77.9	0.0	0.0	2.6	72.2	
	xxv	2.4	3.5	3.5	85.7	3.6	2.5	2.5	34.6	2.5	2.5	77.9	0.0	0.0	2.6	72.2	
	xxvi	2.4	3.5	3.5	85.7	3.6	2.5	2.5	34.6	2.5	2.5	77.9	0.0	0.0	2.6	72.2	

1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457

Table 17: The EEG encoder can generalize in some situations on confounded data with separate training (some of columns iv, viii, xii, xvi, xx, and xxiv, for rows i–iv and ix–xii, are above chance). This, together with Tables 16 and 18, supports claim D.

	before			original			40 classes			bottom 960			top-40			1000 classes			bottom 960			after				
	pretraining	after separate	after	pretraining	after separate	after	pretraining	after separate	after	pretraining	after separate	after	pretraining	after separate	after	pretraining	after separate	after	pretraining	after separate	after	pretraining	after separate	after		
Spanipinato et al. (2017)	i	2.6	3.0	2.2	2.5	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	
	ResNet-101	2.3	3.0	2.8	2.6	2.7	2.4	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	
	ResNet-161	2.3	3.0	2.8	2.6	2.7	2.4	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	
	AlexNet	2.3	2.5	2.4	2.6	2.6	2.4	2.5	2.6	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	
image	v	99.7	43.2	2.7	3.4	99.4	2.9	2.2	45.5	28.4	2.7	2.5	92.8	65.5	0.1	0.0	98.9	4.0	0.0	40.0	12.8	9.3	0.0	0.0	0.0	
	ResNet-101	99.4	24.2	2.4	82.5	98.3	7.3	2.9	70.5	33.0	9.5	2.1	39.4	90.5	2.6	0.0	82.5	95.6	0.3	0.0	70.5	9.5	0.7	0.0	39.2	
	ResNet-161	99.4	24.2	2.4	82.5	98.3	7.3	2.9	70.5	33.0	9.5	2.1	39.4	90.5	2.6	0.0	82.5	95.6	0.3	0.0	70.5	9.5	0.7	0.0	39.2	
	AlexNet	95.2	2.2	2.4	18.2	85.8	2.4	2.7	19.1	34.1	2.6	2.4	28.4	78.4	0.0	0.3	18.7	72.5	0.0	0.3	19.1	12.5	0.0	0.0	2.4	
Tu et al. (2021) block	ix	2.4	6.8	1.9	77.0	1.8	7.1	2.0	69.8	2.6	3.6	2.5	32.1	0.1	0.4	0.0	77.0	0.2	0.1	69.8	0.0	0.2	0.1	25.0	0.0	
	ResNet-101	1.6	3.5	2.2	76.6	1.1	4.2	2.5	70.5	2.4	2.8	2.2	28.1	0.0	0.1	0.1	76.6	0.0	0.1	70.5	0.0	0.0	0.1	24.0	0.0	
	DenseNet-161	1.4	4.1	2.0	77.0	1.2	3.8	1.9	72.2	2.5	3.5	2.7	25.9	0.1	0.1	0.0	77.0	0.0	0.0	72.2	0.1	0.1	0.1	22.3	0.0	
	AlexNet	96.1	3.0	3.0	91.3	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2
image	xi	99.3	75.0	2.4	11.0	98.5	68.5	2.7	3.5	34.1	14.7	2.5	3.7	90.5	56.4	0.0	2.0	96.2	52.4	0.0	0.1	9.8	0.2	0.0	0.0	0.0
	ResNet-101	99.1	95.8	2.5	5.7	97.7	86.8	2.9	3.8	43.9	50.9	2.1	2.6	89.5	71.1	0.1	0.2	96.4	67.2	0.4	0.2	15.0	17.9	0.0	0.0	0.0
	DenseNet-161	95.5	60.9	2.5	3.5	85.7	50.9	2.4	3.2	34.6	15.6	2.5	2.3	77.9	25.4	0.0	2.6	75.2	19.7	0.0	2.5	12.9	4.2	0.0	2.3	
	AlexNet	96.1	3.0	3.0	91.3	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2	91.2
Tu et al. (2021) randomized	xxiii	2.6	3.5	2.8	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6
	ResNet-101	2.1	2.2	2.3	1.5	2.5	2.0	3.1	2.0	1.9	1.9	0.1	0.1	1.9	0.0	0.1	1.9	0.0	0.1	1.5	0.0	0.1	2.0	0.1	0.1	1.7
	DenseNet-161	2.1	2.2	2.3	1.5	2.5	2.0	3.1	2.0	1.9	1.9	0.1	0.1	1.9	0.0	0.1	1.9	0.0	0.1	1.5	0.0	0.1	2.0	0.1	0.1	1.9
	AlexNet	2.5	2.8	3.0	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2
image	xxi	99.4	91.1	3.4	99.4	91.1	3.4	99.4	91.1	3.4	99.4	91.1	3.4	99.4	91.1	3.4	99.4	91.1	3.4	99.4	91.1	3.4	99.4	91.1	3.4	99.4
	ResNet-101	99.3	91.1	3.4	99.4	91.1	3.4	99.4	91.1	3.4	99.4	91.1	3.4	99.4	91.1	3.4	99.4	91.1	3.4	99.4	91.1	3.4	99.4	91.1	3.4	99.4
	DenseNet-161	95.3	65.6	2.1	5.1	97.7	36.0	2.5	3.0	43.9	16.1	2.3	3.0	89.5	30.5	0.1	0.0	96.4	11.7	0.0	0.0	13.6	0.0	0.0	0.0	
	AlexNet	99.5	3.6	3.5	85.7	3.6	3.5	34.6	2.5	2.5	2.5	2.5	77.9	0.0	0.0	2.6	72.2	0.0	2.6	12.9	0.0	2.6	12.9	0.0	2.5	

1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511

Table 18: The EEG encoder cannot generalize on nonconfounded data with joint training with pretraining (all of columns ii, vi, x, xiv, xviii, and xxi, for rows xvii-xx, are at chance). This, together with Tables 16 and 17, supports claim D.

	before		original		bottom 960		top-40		1000 classes		bottom 960		after	
	pretraining	after separate	pretraining	after separate	pretraining	after separate	pretraining	after separate	pretraining	after separate	pretraining	after separate	pretraining	after separate
Spanipinato et al. (2017)	i	2.6	3.0	2.2	2.5	2.3	2.3	0.1	0.1	58.5	0.1	58.5	0.1	2.3
	ii	2.3	3.0	2.4	2.5	2.3	2.3	0.1	0.1	57.4	0.1	57.4	0.1	2.3
	iii	2.3	3.0	2.4	2.5	2.3	2.3	0.1	0.1	57.4	0.1	57.4	0.1	2.3
	iv	2.3	2.5	2.4	2.6	2.5	2.5	0.1	0.1	57.3	0.1	57.3	0.1	2.5
image	v	99.7	43.2	2.7	3.4	99.4	2.7	2.5	28.4	0.1	98.9	0.0	12.8	0.0
	vi	99.4	24.2	2.4	82.5	98.3	7.3	2.9	39.4	90.5	2.6	0.0	82.5	95.6
	vii	99.4	24.2	2.4	82.5	98.3	7.3	2.9	39.4	90.5	2.6	0.0	82.5	95.6
	viii	95.2	2.2	2.4	18.2	85.8	2.4	2.7	78.4	0.0	0.3	18.7	72.5	0.0
EEG	ix	2.4	6.8	1.9	77.0	1.8	7.1	2.0	69.8	2.6	0.1	0.4	0.0	77.0
	x	1.6	3.5	2.2	76.6	1.1	4.2	2.5	70.5	2.4	2.8	2.1	0.1	76.6
	xi	1.4	4.1	2.0	77.0	1.2	3.8	1.9	72.2	2.5	3.5	2.7	0.1	77.0
	xii	99.3	96.1	3.0	93.5	94.2	3.1	2.0	46.0	46.8	2.4	2.8	93.0	78.1
image	xiii	99.3	96.1	3.0	93.5	94.2	3.1	2.0	46.0	46.8	2.4	2.8	93.0	78.1
	xiv	99.3	95.8	2.4	11.0	98.5	68.5	2.7	3.5	34.1	14.7	2.5	3.7	90.5
	xv	99.1	95.8	2.5	5.7	97.7	86.8	2.9	3.8	43.9	50.9	2.1	2.6	89.5
	xvi	95.5	60.9	2.5	3.5	85.7	50.9	2.4	3.2	34.6	15.6	2.5	2.3	77.9
EEG	xvii	2.1	2.2	2.3	2.5	2.0	2.4	2.0	2.4	2.2	2.2	2.2	2.2	2.2
	xviii	2.1	2.2	2.3	2.5	2.0	2.4	2.0	2.4	2.2	2.2	2.2	2.2	2.2
	xix	2.1	2.2	2.3	2.5	2.0	2.4	2.0	2.4	2.2	2.2	2.2	2.2	2.2
	xx	2.5	2.8	3.0	2.2	2.2	2.4	3.0	1.8	0.0	0.1	1.9	0.0	0.1
image	xxi	99.3	91.1	3.4	99.4	99.4	84.7	3.0	2.9	46.0	57.1	2.1	2.5	93.0
	xxii	99.3	91.1	3.4	99.4	99.4	84.7	3.0	2.9	46.0	57.1	2.1	2.5	93.0
	xxiii	95.5	65.6	2.1	5.1	97.7	36.0	2.5	3.0	43.9	36.1	2.3	3.0	89.5
	xxiv	95.5	65.6	2.1	5.1	97.7	36.0	2.5	3.0	43.9	36.1	2.3	3.0	89.5

1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565

Table 19: Classification accuracy is at chance after joint training without pretraining (columns iii vii, xi, xv, xix, and xxiii). This holds both for EEG classification and image classification, both for 40 classes and 1000 classes, both for the ‘Spaminato et al. (2017)’ data, the ‘Li et al. (2021)’ block’ data, and the ‘Li et al. (2021) randomized’ data, and whether using all components, or the bottom 960 components. This suggests that their method completely breaks down when not provided with class information through pretraining, and further supports claim A and calls claim I into question.

		40-classes				bottom 960				top-40				1000-classes												
		before	after separate	original joint	no pretraining joint	before	after separate	original joint	no pretraining joint	before	after separate	original joint	no pretraining joint	before	after separate	original joint	no pretraining joint									
Spaminato et al. (2017)	EEG	i	2.6	2.5	3.0	2.2	2.5	3.1	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3								
		ii	2.5	2.4	2.9	2.2	2.5	3.1	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3								
		iii	2.5	2.4	2.9	2.2	2.5	3.1	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3							
		iv	2.3	2.4	2.6	2.4	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6							
	image	v	99.7	2.7	3.4	99.4	13.5	2.7	2.5	2.7	2.5	92.8	0.1	0.1	0.1	0.1	0.1	0.1	0.1							
		vi	99.4	2.4	2.4	99.4	13.5	2.7	2.5	2.7	2.5	92.8	0.1	0.1	0.1	0.1	0.1	0.1	0.1							
		vii	99.4	2.4	2.4	99.4	13.5	2.7	2.5	2.7	2.5	92.8	0.1	0.1	0.1	0.1	0.1	0.1	0.1							
		viii	95.8	2.6	2.6	85.8	2.6	2.7	19.1	34.1	2.4	78.4	0.1	0.3	18.7	72.5	0.0	0.0	0.0	0.0						
Li et al. (2021) block	EEG	ix	2.4	1.9	6.8	1.8	7.1	2.0	69.8	2.6	3.6	2.5	32.1	0.1	0.1	77.0	0.0	0.2	0.1	68.8	0.0	0.2	0.1	28.0		
		x	1.6	3.5	2.2	76.6	1.1	4.2	2.5	70.5	2.4	2.8	2.2	28.1	0.1	0.1	76.6	0.0	0.1	0.1	70.5	0.0	0.0	0.1	24.0	
		xi	2.0	2.0	2.0	76.0	0.2	3.8	2.2	70.0	2.5	3.4	2.2	27.9	0.1	0.1	77.0	0.0	0.1	0.1	70.0	0.0	0.1	0.1	24.2	
		xii	2.4	2.4	2.4	76.4	2.4	2.4	2.4	76.4	2.4	2.4	2.4	2.4	2.4	2.4	76.4	0.0	0.3	0.2	76.4	0.0	0.1	0.1	24.2	
	image	xiii	97.8	3.0	97.3	97.7	97.7	2.9	46.0	46.8	2.4	2.5	93.0	75.1	0.1	0.5	98.8	87.0	0.0	0.0	100	13.2	11.7	0.0	0.0	
		xiv	99.3	2.4	11.0	98.5	68.5	2.7	3.5	34.1	14.7	2.0	90.5	56.4	0.0	2.0	96.2	52.4	0.0	0.1	9.8	0.2	0.2	0.0	0.0	
		xv	99.1	2.5	5.7	97.7	86.8	2.9	3.8	43.9	50.9	2.1	2.6	89.5	71.1	0.1	0.2	96.4	69.2	0.4	0.2	15.0	17.9	0.0	0.0	0.0
		xvi	92.5	2.5	1.5	85.7	52.9	2.4	2.2	34.6	19.6	2.0	90.5	56.4	0.0	2.0	96.2	52.4	0.0	0.1	9.8	0.2	0.2	0.0	0.0	
Li et al. (2021) randomized	EEG	xvii	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	
		xviii	2.6	2.4	2.8	2.1	2.8	2.7	2.9	2.3	2.6	2.7	3.0	1.8	0.0	0.1	1.9	0.0	0.1	0.1	2.0	0.1	0.2	0.1	1.6	
		xix	2.1	2.2	2.3	1.5	2.5	2.3	2.5	2.0	3.1	2.0	1.9	1.9	0.1	0.1	1.5	0.0	0.0	0.1	2.0	0.1	0.1	0.1	1.7	
		xx	2.5	3.0	2.2	2.2	2.4	2.2	2.4	1.6	2.6	2.8	3.1	2.4	0.2	0.1	2.2	0.3	0.0	0.1	1.6	0.1	0.1	0.1	1.9	
	image	xxi	99.3	2.9	11.2	98.5	53.1	2.7	4.0	47.0	36.0	2.1	2.1	2.1	2.1	2.1	90.5	42.0	0.0	2.0	96.2	28.5	0.0	0.1	9.8	
		xxii	99.3	2.9	11.2	98.5	53.1	2.7	4.0	47.0	36.1	2.1	2.1	2.1	2.1	2.1	90.5	42.0	0.0	2.0	96.2	28.5	0.0	0.1	9.8	
		xxiii	99.1	2.1	5.1	97.7	36.0	2.5	3.0	43.9	36.1	2.3	3.0	89.5	30.5	0.1	0.0	96.4	11.7	0.0	0.0	15.0	13.6	0.0	0.0	
		xxiv	95.5	3.5	85.7	3.6	2.5	2.5	77.9	0.0	0.0	2.6	72.2	0.0	0.0	0.0	2.6	72.2	0.0	0.0	2.6	12.9	0.0	0.0	2.5	

1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619

Table 20: With only two exceptions, separate training gets to a lower loss than joint training with pretraining (compare **columns vii, viii, and ix** to **i, ii, and iii**, respectively). This suggests that there is a point within the representational-capacity space of their model and loss function that achieves lower loss than achieved by their joint-training procedure. The joint-training procedure could have achieved it, it just didn't. Since that point was achieved with separate training, the resulting EEG encodings do not have any image information and the resulting image encodings do not have any brain-activity information. This supports claim B and calls claims II-IV into question.

		joint						separate		
		pretraining			no pretraining			Spampinato et al. (2017)	Li et al. (2021) block	Li et al. (2021) randomized
		i	ii	iii	iv	v	vi	vii	viii	ix
Inception v3	i	1.303	1.098	0.932	1.074	0.938	0.919	1.150	1.027	1.104
ResNet-101	ii	1.155	0.925	1.019	0.963	1.006	1.080	0.885	0.936	0.884
DenseNet-161	iii	1.027	1.080	1.047	0.988	0.919	0.929	1.007	0.982	1.029
AlexNet	iv	1.417	1.142	0.981	1.136	1.055	1.034	0.925	0.956	0.935

Table 21: With only three exceptions, joint training without pretraining gets to a lower loss than with pretraining (compare **columns iv, v, and vi** to **i, ii, and iii**, respectively). Yet classification accuracy is at chance after joint training without pretraining (columns iii, vii, xi, xv, xix, and xxiii of Table 2). This suggests that joint training without pretraining can memorize the training set yet fail to generalize at all, and further supports claim A and calls claim I into question.

		joint						separate		
		pretraining			no pretraining			Spampinato et al. (2017)	Li et al. (2021) block	Li et al. (2021) randomized
		i	ii	iii	iv	v	vi	vii	viii	ix
Inception v3	i	1.303	1.098	0.932	1.074	0.938	0.919	1.150	1.027	1.104
ResNet-101	ii	1.155	0.925	1.019	0.963	1.006	1.080	0.885	0.936	0.884
DenseNet-161	iii	1.027	1.080	1.047	0.988	0.919	0.929	1.007	0.982	1.029
AlexNet	iv	1.417	1.142	0.981	1.136	1.055	1.034	0.925	0.956	0.935