Continual Learning of Object Instances

Anonymous Authors[1]

Abstract

We propose continual instance learning - a method that applies the concept of continual learning to the task of distinguishing instances of the same object category. We specifically focus on the car object, and incrementally learn to distinguish car instances from each other with metric learning. We begin our paper by evaluating current techniques. Establishing that catastrophic forgetting is evident in existing methods, we then propose two remedies. Firstly, we regularise metric learning via Normalised Cross-Entropy. Secondly, we augment existing models with synthetic data transfer. Our extensive experiments on three large-scale datasets, using two different architectures for five different continual learning methods, reveal that Normalised cross-entropy and synthetic transfer leads to less forgetting in existing techniques.

1. Introduction

Most computer vision tasks assume access to a full, static dataset. Regularly, researchers train and test their algorithms on well-established benchmarks (Krizhevsky et al., 2012; Everingham et al., 2010; Lin et al.). Although beneficial from a benchmarking point of view, this setup neglects the dynamic, ever-changing nature of the visual world. The world does not present itself as a static set of objects that remain similar through time. Humans evolved to be life-long learners (Mcclelland et al., 1995) and update their visual model of the world with sensory data. Therefore, continual learning (CL) is proposed to mimic learning about novel object categories (Li & Hoiem, 2016; Jung et al., 2016; Kirkpatrick et al., 2017).

A common task for CL is to learn about a new object (*e.g.* chair) while retaining performance on the previously seen (*e.g.* car, table, bicycle). In this work, we focus on *continuously* learning to distinguish between different instances of the same object category, in our case, cars. Continual instance learning (CIL) is an approach that applies the concept of CL to the task of distinguishing intra-category instances through *metric learning* (Kuma et al., 2019; Schroff et al., 2015; Chopra et al., 2005).

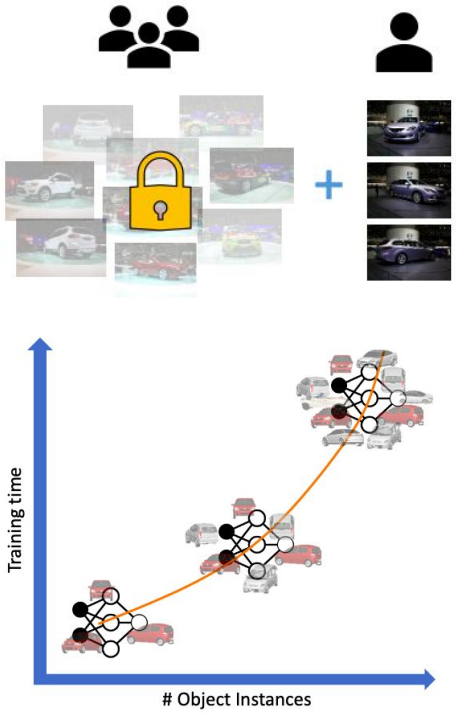

Figure 1. Two cases for continual learning of object instances. (Top) Retail companies receive thousands of novel car instances daily. However, these are not always accessible due to privacy reasons. (Bottom) Surveillance companies receive long streams of car records, therefore training from scratch becomes inefficient. In this paper, we propose continual learning of object instances to tackle such issues of privacy and inefficiency.

A continuous stream of car objects is evident in the visual world, especially for cases of retail and surveillance, consider Figure 1. For instance, an online car retail company receives numerous new car advertisements daily. Ideally, a company would continuously learn from the incoming source of car images, for applications such as identifying

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

duplicate ads (Zhang et al., 2018). However, due to privacy reasons, companies may need to delete historical user records, which affects model (re-)training (Goodfellow et al., 2013; Lomonaco & Maltoni, 2017). Likewise, a surveillance company may aim to learn to re-identify a query vehicle within the city. However, learning vehicle re-identification from scratch is not optimal, since the amount of car images is ever-increasing. We address such scenarios with continual instance learning.

Our contributions are as follows:

1. We introduce the problem of *continual instance learning* by evaluating existing continual learning methods under the metric learning setup. We show that existing techniques suffer from *catastrophic forgetting*.

2. We propose to utilise Normalised Cross-Entropy (NCE) (Chen et al., 2020), to reduce the effects of catastrophic forgetting by mitigating the sensitivity to outliers under regression losses (Hermans et al., 2017; Schroff et al., 2015).

3. We augment existing techniques with synthetic visual data transfer, showing improvements on three benchmark datasets over two backbone architectures.

2. Related Work

2.1. Continual Learning

Under different training scenarios, neural networks suffer from catastrophic forgetting (McCloskey & Cohen, 1989; Goodfellow et al., 2013; Lomonaco & Maltoni, 2017). This pitfall has inspired the research community to reestablish neural networks as a system that is not only put on hold once trained but as one that can improve through time. This has given rise to continual learning, also known as incremental learning, formally introduced in (Lomonaco & Maltoni, 2017). In continual learning, one does not have access to previously seen data.

Our work follows more closely the New Instances (NI) setting (Lomonaco & Maltoni, 2017) - new training data of the same classes become available with new poses and conditions. We differentiate ourselves by presenting new instances of only a single class object and learn to better distinguish these. We focus on regularisation approaches since these have shown to have adequate performances with a low level of complexity (Jung et al., 2016; Li & Hoiem, 2016; Kirkpatrick et al., 2017). Early attempts regularise the loss function by maintaining the output of the network as unchanged as possible while shifting the internal feature representation, namely Less Forgetting Learning (LFL) (Jung et al., 2016) and Learning without Forgetting (LwF) (Li & Hoiem, 2016). In Elastic Weight Consolidation (EWC) importance is defined for each model parameter via a Bayesian approach (Kirkpatrick et al., 2017). This research is closely related to (Lomonaco & Maltoni, 2017) and (Stojanov et al., 2019) where we differentiate ourselves from the classification setup and introduce metric learning to continual learning.

2.2. Instance Learning

In metric learning one directly learns a distance function between objects in a D-dimensional space. Applications in this domain extend to image retrieval tasks, specifically, face verification (Chopra et al., 2005), person and vehicle re-identification (ReID) (Schroff et al., 2015; Kuma et al., 2019). In this work, we focus on the vehicle re-identification problem. Common vehicle ReID learning approaches resort to Siamese CNN (Bromley et al., 1993) with contrastive (Hadsell et al., 2006) or triplet losses (Schroff et al., 2015; Kuma et al., 2019). These introduce an effective method of separating objects in the embedding space.

Directly translating CL approaches to metric learning is not possible due to the different problem formulation - these approaches build upon regularising classification outputs, whereas metric learning directly attempts to learn a manifold. This approach is susceptible to outliers (Schroff et al., 2015) due to the unbounded nature of the gradients. To this end, we propose Normalised Cross-Entropy (Chen et al., 2020) as a solution. More specifically, we treat metric learning as a classification task, where given a query image, we solve a binary task to estimate whether a retrieved car image belongs to the same query under a different viewpoint.

2.3. Transfer Learning from Synthetic Data

The successful effect of using synthetic data for improving machine learning systems has been reported in (Gaidon et al., 2018). It reduces the need for labels (Paetzold et al., 2019) and has shown to help in different visual tasks (Rogez & Schmid, 2018; Sakaridis et al., 2017). To reduce forgetting, we evaluate the effect of pre-training a model on object-specific discriminant features. For this, we propose to utilise synthetic transfer learning for CIL.

3. Approach

3.1. Continual Learning

Regularisation CL approaches are easily adapted to systems in production. These do not rely on historical samples and do not increase the number of model parameters. In our experiments, we focus on Less Forgetting Learning (Jung et al., 2016), Learning without Forgetting (Li & Hoiem, 2016) and Elastic Weight Consolidation (Kirkpatrick et al., 2017). The first two approaches focus on keeping the deci-

sion boundary as unchanged as possible, whereas the latter targets updating internal representations. For comparison, we also run our experiments on baseline CL approaches, specifically, Naïve and Fine-tuning.

We define X_o and Y_o as the input and targets of an *old* dataset that is no longer available. Similarly, for a *new* dataset, we define X_n and Y_n. Additionally, the parameters of a previously trained model and the new model are defined as θ_o and θ_n respectively.

Naïve: The benchmark approach for continual learning is to simply re-train a model on X_n and Y_n given θ_o. This training procedure does not require any additional parameters and by definition, it yields θ_n.

Fine-Tuning (FT): One simple approach to bypass catastrophic forgetting is to first train a model on X_o and Y_o, and once we obtain access to X_n and Y_n, train a model by fine-tuning discriminant layers (Li & Hoiem, 2016).

Less Forgetting Learning (LFL) (Jung et al., 2016): One important characteristic in CL is to ensure that we get similar predictions $\hat{Y}$ for X_n under θ_o and θ_n. To this end, in (Jung et al., 2016) the authors propose to improve the re-training process by initialising a model with $\theta_n = \theta_o$. Additionally, it is proposed to freeze the softmax layer so that the decision boundaries of the model remain similar. The authors train such a model with the following loss function:

$$\mathcal{L}_{LFL}(x_n; \theta_o; \theta_n) = \lambda_c \mathcal{L}_c(x_n; \theta_n) + \lambda_e \mathcal{L}_e(x_n; \theta_o; \theta_n), \tag{1}$$

where $\mathcal{L}_c$ and $\mathcal{L}_e$ are the cross-entropy and the Euclidean loss functions. The authors constrain feature changes with the Euclidean loss between the predictions of X_o under θ_n and X_o under θ_o.

Learning without Forgetting (LwF) (Li & Hoiem, 2016): Similarly, in LwF (Li & Hoiem, 2016), the authors encourage the network to keep its features as unchanged as possible by using the Knowledge Distillation loss (Hinton et al., 2014) as opposed to the Euclidean loss in Equation 1. However, in this approach, θ_n is randomly initialised and there is no explicit network freezing.

Elastic Weight Consolidation (EWC) (Kirkpatrick et al., 2017): Restricting feature extraction capabilities is not optimal since we want a model to learn from new instances and incrementally improve its object representation. Therefore, in (Kirkpatrick et al., 2017) the model parameters θ take a probabilistic form with the Fisher Information matrix, F_i, reflecting weight importance. Under CL, the model is regularised with the following loss function:

$$\mathcal{L}_{EWC} = \mathcal{L}_n + \sum_i \frac{\lambda}{2} F_i(\theta_{n,i} - \theta_{o,i}), \tag{2}$$

where $\mathcal{L}_n$ is the loss function for the new task, and λ is the previous task importance.

3.2. Regression-based Metric Learning

We train the vehicle ReID datasets using a siamese network and the triplet loss (Schroff et al., 2015). We distance dissimilar vehicles and cluster similar ones in the manifold space. Specifically, we define an anchor (x_i^a), a positive (x_i^p) and a negative sample (x_i^n). The loss function is defined as:

$$\sum_i^N \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+, \tag{3}$$

where α is defined as the margin which, similar to (Hermans et al., 2017), is set to 1. *Mining* hard negative pairs is highly important to obtain discriminant features *i.e.*, we want to emphasise training of instances which are harder to distinguish, such as two different cars of similar colour and pose, as opposed to trivially distinct cars.

4. Continual Instance Learning

4.1. Continuous Batches

We define three data streaming approaches for accessing object instances under CIL, where one time-step, or a new training set, is a *continuous batch*. Each approach presents different restrictions and is suitable for different applications.

1. **Random:** Upon acquiring new data, the particular object instances in the set are not known *a priori*. Therefore, this approach assumes that in each continuous batch there can be different instances and each of them can have a different number of data-points. Under the triplet loss, we do not know the number of negative instances, or how hard these are to learn from.
2. **Balanced:** Some tasks can also acquire new instance data in a balanced fashion *i.e.*, the same number of data-points are collected in each time step for all instances. If this is the case, we hypothesize that it is easier to mine for hard negatives, since we have access to more negative instances.
3. **Incremental:** On the other hand, some applications acquire only some instances at a time *i.e.*, every continuous batch is restricted to only a few instances. This approach has limited access to negative instances and is, therefore, the most challenging.

In our experiments, we assume the most challenging scenario, incremental continuous batches, which is also most closely related to the examples portrayed in section 1.

4.2. Adapting current approaches

We adapt CL approaches to CIL for bench-marking purposes. To achieve this, we propose the following changes:

- For the Naïve and Fine-Tuning approaches, we directly replace the cross-entropy loss (Li & Hoiem, 2016) with the triplet loss.
- For LFL and EWC we propose the same approach by replacing $\mathcal{L}_c$ from Equation 1 and $\mathcal{L}_n$ from Equation 2 with the triplet loss.
- In the LwF scenario, the authors make use of the knowledge distillation loss, which encourages probabilities of one network to approximate the output of another network. Since in the regression-based triplet loss we are not dealing with probabilities, we are therefore not able to adapt this approach.

4.3. Normalised Cross-Entropy Loss (NCE)

Two unwanted features in continual instance learning are, 1) dealing with a regression setup which is not robust to outliers and 2) not being able to build upon current CL approaches, like LwF.

Having an outliers-robust approach allows us to have instances that are extremely different from all others and still restrict its gradients. This can be a fairly common scenario in incremental continuous batches. To this end, the triplet loss is adapted to make use of the normalised cross-entropy loss (Chen et al., 2020),

$$\ell(z_i) = -\log \frac{\exp(z_i/\tau)}{\sum_{j=1}^{K} \exp(z_j/\tau)} \text{ for } i = 1, \ldots, K+1 \quad (4)$$

where,

$$\mathbf{z} = \left(x^{a\top} x^{p}, x^{a\top} x_1^{n} \ldots, x^{a\top} x_K^{n} \right). \quad (5)$$

We set the temperature, $\tau = 1$ for simplicity. We do not resort to offline or online negative mining since we are constrained by the number of hard negative samples. However, we allow for the loss function to include more than one negative sample at a time (Sohn, 2016). We define the target of $\mathbf{z}$ as 1 for the index corresponding to the dot product between the anchor and the positive pair and 0 for all other pairs

4.4. Synthetic Visual Data Transfer

We first train a model on synthetic instances of the same object as the real object of interest. We then apply CIL under new real instances. We motivate this approach to ensure that the model first fully learns fine-grained visual representations from the synthetic data before training on continuous batches, which may be of skewed distributions.

5. Experimental Setup

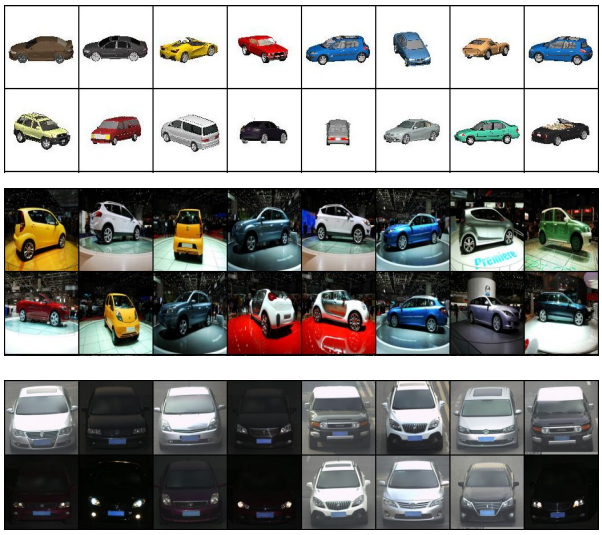

Figure 2. Example images from the three datasets, namely Cars3D (top), MVCD (middle) and CompCars (bottom). Cars3D is synthetic and other two datasets are real datasets.

5.1. Datasets

We use three publicly available datasets throughout our experiments. For each dataset, we split the training data in a continuous incremental batch fashion. We do not make use of any dataset-specific procedure (*i.e.*, augmentation) in our experiments.

Cars3D (Reed et al., 2015) Consists of 183 car models, for each model we render 96 data points generated from 24 equally spaced azimuth directions and 4 elevations. We use 100 car models for training. For evaluating, we use 83 car models. From each, we randomly select 10 images for the query set and 86 images for the gallery set. In total, the training set consists of 9600 images, split into 10 incremental batches. The query and gallery set consist of 830 and 6972 images respectively.

Multi-View Car Dataset (Ozuysal et al., 2009) (MVCD) Consists of a sequence of images from 20 different car models on a rotating platform. An image was taken every 3 to 4 degrees. For training, we select 15 car models, comprising of 1737 images split into 5 incremental batches. From the remaining 5 car models, we select 10 images for the query set and the remaining for the gallery set.

Table 1. Benchmarking existing techniques.

		LeNet			ResNet		
Dataset	Approach	Ref(%)	mAP(%↑)	Forget(%↓)	Ref(%)	mAP(%↑)	Forget(%↓)
Cars3D	Naïve	68.69	46.28	32.89	70.82	50.01	29.38
	FT	68.69	**49.36**	**28.43**	70.82	44.98	36.49
	LFL	68.69	33.05	52.08	70.82	33.96	52.05
	LwF	68.69	–	–	70.82	–	–
	EWC	68.69	46.19	33.02	70.82	**50.91**	**28.11**
MVCD	Naïve	83.22	62.31	25.13	94.99	**83.77**	**11.81**
	FT	83.22	60.65	27.13	94.99	72.75	23.41
	LFL	83.22	**62.86**	**24.47**	94.99	66.24	30.27
	LwF	83.22	–	–	94.99	–	–
	EwC	83.22	61.44	26.18	94.99	81.44	14.26
CompCars	Naïve	29.83	**18.25**	**38.82**	54.41	**38.60**	**29.06**
	FT	29.83	12.51	58.06	54.41	22.63	58.41
	LFL	29.83	11.18	62.52	54.41	25.24	53.61
	LwF	29.83	–	–	54.41	–	–
	EWC	29.83	17.19	42.37	54.41	37.43	31.21

CompCars (Yang et al., 2015) We use the surveillance-nature subset which contains 50,000 car images, captured only from the frontal view. In total, there are 281 different car models. The number of images for each model ranges from 50 to 500 different captures. We use 240 car models for our training set, a total of 36737 images split into 10 incremental batches. For the query set, we randomly select 20 images of the unseen cars and use the remaining images as the gallery set.

5.2. Backbone Architectures

We use two different backbone architectures - both trained from scratch. All models are trained until convergence. We first train a model with the first continuous batch and then re-train the same model with subsequent continuous batches and a continual instance learning approach. We repeat this process until all instances are trained.

Unless specifically stated, we use the same training procedure across all experiments and datasets. We implement our experiments using the PyTorch framework (Paszke et al., 2017). We use the Adam optimizer (Kingma & Ba, 2014) with $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-3}$ for the LeNet experiments and $\epsilon = 10^{-4}$ for the ResNet experiments.

LeNet - We use a LeNet-5 (LeCun et al., 1998) like architecture with 3 convolutional layers and 3 fully connected layers. We apply batch normalisation (Ioffe & Szegedy, 2015), ReLU non-linearities and 2 x 2 max-poolings with stride 2. All layers are initialised with Xavier uniform (Glorot & Bengio, 2010). For all datasets, we train the network with the 32 x 32 RGB images and define the last fully connected layer with D-dimension 32.

ResNet - We employ the ResNet-18 (He et al., 2016) with 64 x 64 RGB images for the Cars3D (Reed et al., 2015) and the Multi-View Car Dataset (Ozuysal et al., 2009), and 128 x 128 RGB images for the CompCars dataset (Yang et al., 2015). We use the last fully connected layer as the metric extractor with dimension 128.

5.3. Evaluation.

In the re-identification evaluation setup, we have a query set and a gallery set. For each vehicle in the query set, we aim to retrieve the same vehicle instances from the gallery set. We use mean-average-precision (mAP) for evaluating and comparing our approaches. The mAP is computed as the Average Precision (AP) across all queries.

Under the continual learning framework, we employ a similar approach as (Lomonaco & Maltoni, 2017) and compare the relative mAP w.r.t. the corresponding cumulative approach - training with all instances - which is hereby referred to as *Forget*. Our metrics on a Full Test Set: the query and gallery sets remain fixed throughout the continuous batches.

6. Experiments

Our experiments aim at answering the following questions **Q1)** What is the performance of existing continual instance

Table 2. Contribution of Normalised Cross Entropy (NCE) to Continual Instance Learning - Forget ratio.

		LeNet		ResNet	
Dataset	Approach	Regression(%↓)	w/ NCE(%↓)	Regression(%↓)	w/ NCE(%↓)
Cars3D	Naïve	32.89	25.64	29.38	30.61
	FT	28.43	24.02	36.49	33.71
	LFL	52.08	33.50	52.05	43.10
	LwF	–	**18.29**	–	30.19
	EWC	33.02	23.66	28.11	**19.87**
MVCD	Naïve	25.13	22.80	**11.81**	16.10
	FT	27.13	**20.37**	23.41	24.16
	LFL	24.47	20.76	30.27	27.11
	LwF	–	21.36	–	14.59
	EWC	26.18	28.44	14.26	13.92
CompCars	Naïve	38.82	35.13	**29.06**	39.77
	FT	58.06	51.66	58.41	52.38
	LFL	62.52	58.26	53.61	54.47
	LwF	–	49.55	–	60.37
	EWC	42.37	**32.28**	31.21	29.13

learning methods for object instance recognition? **Q2)** Can Normalised Cross-Entropy improve CIL? **Q3)** Can synthetic transfer learning improve CIL?

6.1. Q1: Performance of Existing Techniques

In Table 1, we evaluate the different continual learning approaches under the CIL setup. Additionally, we provide the cumulative mAP as an offline reference. Similar to the continual classification task, CIL also suffers from catastrophic forgetting.

In general, the Naïve approach outperforms others. Additionally, EWC approach also has good performance. In both approaches, the network is allowed to fully improve its decision boundary and its feature extraction capacity. The LFL approach has, in general, the lowest performance. We hypothesize that keeping the embedding unchanged is not desirable when we aim to learn the embedding itself. The ResNet-18 model performs better under the cumulative approach and suffers from less forgetting when compared to the LeNet model. The CompCars dataset is the most challenging set under offline and CIL, whereas the MVCD is the easiest.

In this experiment, we conclude that Continual Instance Learning is a challenging problem for current CL approaches, and the existing methods perform comparably or inferior to the Naïve approach.

6.2. Q2: Contribution of NCE

In our second experiment, we investigate the effect of NCE, Table 2. In all experiments, we notice that the more negatives samples we add, the better the results. We restrict to nine negative samples.

We identify that most methods benefit from NCE. This is particularly true for EWC, where the feature extracting layers benefit the most from bounding the gradients. Overall, both architectures gain from NCE. This is more visible in the LeNet model which is more prone to over-fitting and more highly impacted by regression outliers. Integrating NCE to CIL is straightforward and beneficial to most approaches.

6.3. Q3: Contribution of Synthetic Transfer

In Table 3, we investigate the effects of pre-training a network on synthetic data on CIL approaches. Synthetic object instances can be found in publicly available datasets (Chang et al., 2015; Xiang et al., 2016). Specifically, in our experiments, we pre-train a model on the entire Cars3D dataset and perform continuous batch training on MVCD and CompCars. We reduced the learning rate to 10^{-4} for all continual experiments and froze the convolutional layers.

All methods, under both architectures, benefit from our method. Particularly, Naïve and EWC. This is more evident in MVCD. The gains are not as visible in CompCars. This is likely due to the differences in the data distributions. The synthetic dataset consists of car images in different azimuths and elevations. Conversely, CompCars has only frontal view

Table 3. Contribution of Synthetic Transfer to Continual Instance Learning - Forget ratio.

		LeNet		ResNet	
Dataset	Approach	Regression(% ↓)	w/ NCE+Transfer(%↓)	Regression(% ↓)	w/ NCE+Transfer(%↓)
MVCD	Naïve	25.13	9.84	11.81	**0.48**
	FT	27.13	10.15	23.41	19.23
	LFL	24.47	14.25	30.27	13.16
	LwF	–	**3.08**	–	12.12
	EWC	26.18	9.22	14.26	2.29
CompCars	Naïve	38.82	**37.85**	**29.06**	33.82
	FT	58.06	71.74	58.41	70.99
	LFL	62.52	61.28	53.61	50.01
	LwF	–	60.44	–	41.15
	EWC	42.37	39.73	31.21	31.35

images. This restricts the visual cues to features, which are limited in Cars3D, such as the brand, the frontal grill, and the headlamps. Overall, the ResNet model benefits the most from synthetic transfer. Having a relevant distributed synthetic dataset can reduce forgetting in CIL in combination with NCE.

7. Conclusion

In this paper, we studied continual learning of object instances. Firstly, we verified that existing techniques suffer from catastrophic forgetting while learning object instances. We show that the simple Naïve approach performs competitively to other continual learning techniques. This indicates that existing methods are unsuitable for continual instance learning, calling for specific techniques. To that end, we incorporated normalised cross-entropy along with synthetic visual transfer to existing techniques to circumvent forgetting. We observed that indeed cross-entropy regulates the learning, and synthetic visual data is beneficial, especially when it is similar to the target data. From our observations, we foresee two plausible directions for future research. In this paper, we solely build upon regularisation approaches, however, replay-based techniques could be promising for continual instance learning. Secondly, our paper focused on car instances, and we leave the exploration of other rigid object instances, such as household objects, or nonrigid object instances such as humans or human faces as future work.

References

Bromley, J., Bentz, J., Bottou, L., Guyon, I., Lecun, Y., Moore, C., Sackinger, E., and Shah, R. Signature verification using a "siamese" time delay neural network. *IJPRAI*, 1993.

Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., and Yu, F. ShapeNet: An Information-Rich 3D Model Repository. Technical Report arXiv:1512.03012, 2015.

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. A simple framework for contrastive learning of visual representations. *arXiv preprint arXiv:2002.05709*, 2020.

Chopra, S., Hadsell, R., and LeCun, Y. Learning a similarity metric discriminatively, with application to face verification. In *CVPR*, 2005.

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. The pascal visual object classes (voc) challenge. *IJCV*, 2010.

Gaidon, A., López, A., and Perronnin, F. The reasonable effectiveness of synthetic visual data. In *IJCV*, 2018.

Glorot, X. and Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In *AISTATS*, 2010.

Goodfellow, I. J., Mirza, M., Xiao, D., Courville, A., and Bengio, Y. An empirical investigation of catastrophic forgetting in gradient-based neural networks. In *ICLR*, 2013.

Hadsell, R., Chopra, S., and LeCun, Y. Dimensionality reduction by learning an invariant mapping. In *CVPR*, 2006.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *CVPR*, 2016.

Hermans, A., Beyer, L., and Leibe, B. In defense of the triplet loss for person re-identification. In *CVPR*, 2017.

Hinton, G., Vinyals, O., and Dean, J. Distilling the knowledge in a neural network. In *NIPS*, 2014.

Ioffe, S. and Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *JMLR*, 2015.

Jung, H., Ju, J., Jung, M., and Kim, J. Less-forgetting learning in deep neural networks. *arXiv*, 2016.

Kingma, D. and Ba, J. Adam: A method for stochastic optimization. In *ICLR*, 2014.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al. Overcoming catastrophic forgetting in neural networks. 2017.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.

Kuma, R., Weill, E., Aghdasi, F., and Sriram, P. Vehicle re-identification: an efficient baseline using triplet embedding. In *IJCNN*, 2019.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998.

Li, Z. and Hoiem, D. Learning without forgetting. In *ECCV*, 2016.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. Microsoft coco: Common objects in context. In *ECCV*.

Lomonaco, V. and Maltoni, D. Core50: a new dataset and benchmark for continuous object recognition. *PMLR*, 2017.

Mcclelland, J., Mcnaughton, B., and O'Reilly, R. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, 1995.

McCloskey, M. and Cohen, N. J. Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of learning and motivation*. 1989.

Ozuysal, M., Lepetit, V., and Fua, P. Pose estimation for category specific multiview object localization. In *CVPR*, 2009.

Paetzold, J. C., Schoppe, O., Al-Maskari, R., Tetteh, G., Efremov, V., Todorov, M. I., Cai, R., Mai, H., Rong, Z., Ertuerk, A., et al. Transfer learning from synthetic data reduces need for labels to segment brain vasculature and neural pathways in 3d. In *MIDL*, 2019.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. Automatic differentiation in pytorch. In *NIPS*, 2017.

Reed, S. E., Zhang, Y., Zhang, Y., and Lee, H. Deep visual analogy-making. In *NIPS*. 2015.

Rogez, G. and Schmid, C. Image-based synthesis for deep 3d human pose estimation. In *IJCV*, 2018.

Sakaridis, C., Dai, D., and Van Gool, L. Semantic foggy scene understanding with synthetic data. In *IJCV*, 2017.

Schroff, F., Kalenichenko, D., and Philbin, J. Facenet: A unified embedding for face recognition and clustering. In *CVPR*, 2015.

Sohn, K. Improved deep metric learning with multi-class n-pair loss objective. In *NIPS*, 2016.

Stojanov, S., Mishra, S., Thai, N. A., Dhanda, N., Humayun, A., Yu, C., Smith, L. B., and Rehg, J. M. Incremental object learning from contiguous views. In *CVPR*, June 2019.

Xiang, Y., Kim, W., Chen, W., Ji, J., Choy, C., Su, H., Mottaghi, R., Guibas, L., and Savarese, S. Objectnet3d: A large scale database for 3d object recognition. In *ECCV*, 2016.

Yang, L., Luo, P., Loy, C. C., and Tang, X. A large-scale car dataset for fine-grained categorization and verification. In *CVPR*, 2015.

Zhang, Y., Zhang, Y., Sun, J., Li, H., and Zhu, Y. Learning near duplicate image pairs using convolutional neural networks. 2018.