GENERATIVE HINTS

Dimnaku, Andy*
Department of Electrical Engineering
Stanford University
Palo Alto, CA 94305, USA
adimnaku@stanford.edu

Kavranoglu, A. Yusuf
Department of Electrical Engineering
California Institute of Technology
Pasadena, CA 91125, USA
ykavranoglu@gmail.com

Abu-Mostafa, Yaser
Department of Electrical Engineering
California Institute of Technology
Pasadena, CA 91125, USA
yaser@caltech.edu

ABSTRACT

Data augmentation is widely used in vision to introduce variation and mitigate overfitting, by enabling models to learn invariant properties. However, augmentation only indirectly captures these properties and does not explicitly constrain the learned function to satisfy them beyond the empirical training set. We propose *generative hints*, a training methodology that directly enforces known functional invariances over the input distribution. Our approach leverages a generative model trained on the training data to approximate the input distribution and to produce *unlabeled* synthetic images, which we refer to as *virtual examples*. On these virtual examples, we impose hint objectives that explicitly constrain the model's predictions to satisfy known invariance properties, such as spatial invariance. Although the original training dataset is fully labeled, generative hints train the model in a semi-supervised manner by combining the standard classification objective on real data with an auxiliary hint objectives applied to unlabeled virtual examples. Across multiple datasets, architectures, invariance types, and loss functions, generative hints consistently outperform standard data augmentation, achieving accuracy improvements of up to 2.10% on fine-grained visual classification benchmarks and an average gain of 1.29% on the CheXpert medical imaging dataset.

1 INTRODUCTION

Data augmentation has become ubiquitous in computer vision, with transformations such as rotations, crops, and color jittering serving as essential components of modern training pipelines (Shorten & Khoshgoftaar, 2019). By exposing models to transformed versions of training examples, augmentation helps models implicitly learn invariance to these transformations. For instance, by training on both an original image and its horizontally flipped version with the same label, a classifier learns that predictions should remain unchanged under horizontal flips. This implicit learning of invariances has proven crucial for achieving state-of-the-art performance across diverse vision tasks (Perez & Wang, 2017).

Despite its widespread success, data augmentation has two fundamental limitations. First, it only implicitly encourages invariance through exposure to transformed examples, without explicitly enforcing that the model's predictions satisfy invariance constraints. The model may learn to be invariant on augmented training examples while still violating invariance properties on unseen data from the input distribution. Second, augmentation is inherently limited to the finite training set, whereas the true input distribution is continuous and extends beyond observed examples. These

*Work done while at California Institute of Technology.

limitations motivate the need for approaches that can explicitly enforce invariances over the learned input distribution.

We build on the concept of *hints* from (Abu-Mostafa, 1990), which encode known properties of the target function as auxiliary training objectives. In the original formulation for tabular data, hints were enforced on random samples from simple noise distributions (e.g., uniform or Gaussian noise), which sufficiently covered the low-dimensional input space. However, in high-dimensional vision tasks, random noise lies far from the natural image manifold, rendering this strategy ineffective. To address this, we leverage modern generative models to approximate the input distribution and sample realistic synthetic images as *virtual examples* (following the terminology of (Abu-Mostafa, 1995)) on which to enforce hints.

Our approach introduces an auxiliary training objective on synthetic data: while the classification loss operates on labeled training examples, the hint loss operates on unlabeled virtual examples sampled from a generative model trained on the training data. For instance, to enforce spatial invariance, we sample a virtual example x_v from the generative model, apply a spatial transformation $h(x_v)$ (such as rotation or translation), and minimize the divergence between the model's predictions on x_v and $h(x_v)$. This explicitly enforces that the model's learned function satisfies $f(x_v) = f(h(x_v))$ over the distribution approximated by the generative model, rather than just on the finite training set used by standard augmentation.

Crucially, generative hints compound with standard data augmentation rather than replacing it. Both our baseline and hint-based training use identical augmentation strategies on labeled training data; the only difference is that hint-based training additionally enforces invariance constraints on unlabeled virtual examples. This controlled design allows us to isolate the contribution of explicit invariance enforcement. We demonstrate that this combination consistently outperforms augmentation-only baselines across diverse settings. Specifically, Our key contributions are:

1. We demonstrate that generative hints consistently improve over data augmentation baselines across multiple architectures (ViT-B, Swin-B, ResNet50), datasets (Stanford Cars, FGVC Aircraft, CUB-200-2011, Oxford Flowers, CheXpert), and loss functions (cross-entropy and MSE). Hints achieve accuracy gains of up to 2.10% on fine-grained visual classification and an average improvement of 1.29% on CheXpert medical imaging.
2. To our knowledge, we are the first to reformulate a fully supervised visual classification task with labeled training data as a semi-supervised training problem by treating data synthesized from a generative model as *unlabeled data*.
3. We introduce a training methodology that explicitly enforces known invariance properties of the target function over the learned input distribution, rather than relying solely on implicit learning through data augmentation on the finite training set.

2 Related Work

2.1 Generative Models for Vision

Generative Models Recent advances in generative modeling have enabled the synthesis of high-fidelity images from noise, primarily through diffusion models (Ho et al., 2020; Song et al., 2021; Rombach et al., 2022) and GANs (Goodfellow et al., 2014; Karras et al., 2019; 2020b; 2021). These models have been applied both as tools for data generation and as components of downstream training pipelines. In classification, discriminators have been adapted for semi-supervised learning (Kingma et al., 2014; Radford et al., 2016), while synthetic data has been used to expand training sets in medical and natural image domains (Antoniou et al., 2017; Frid-Adar et al., 2018; Zhao et al., 2019; Azizi et al., 2023; Yuan et al., 2024). More recently, diffusion-based pipelines (Bordes et al., 2023; Huang et al., 2023; Zhang et al., 2024) highlight the ability of generative models to provide controllable, task-aware augmentation. However, these approaches primarily focus on increasing data diversity rather than directly enforcing functional properties.

Data Augmentation and Invariances Conventional data augmentation is widely used to induce invariances (e.g., spatial or color invariance) by perturbing training examples. While effective for regularization, this strategy only encourages models to learn invariance indirectly, relying on the

hope that augmented samples approximate invariance-preserving transformations (Perez & Wang, 2017; Shorten & Khoshgoftaar, 2019).

Generative Data Augmentation Generative data augmentation (GDA) builds on generative models to synthesize additional labeled data, with demonstrated benefits in low-data regimes, domain-specific applications, and joint generation–classification frameworks (Mahapatra & Ge, 2022). Yet, existing GDA methods treat generated examples primarily as extra training data, without using them to explicitly encode known invariances or functional constraints.

Unlike standard augmentation or GDA, generative hints use synthetic examples as unlabeled carriers of functional properties. That is, generative hints focus is on learning properties of the target function through our semi-supervised training so it can be additively done with existing GDA works.

2.2 HINTS

Hints were first introduced in (Abu-Mostafa, 1990; 1995) to teach machine learning models functional properties of the target function and data. These properties, referred to as hints, are incorporated as auxiliary objectives optimized alongside the main task. For example, in credit default prediction using tabular data, the target is to predict whether a default will occur given input features. Domain knowledge provides that, if all other features remain fixed while income increases, the probability of default should decrease. This property can be formalized as a *monotonicity hint* and enforced through an auxiliary loss. Similarly, in the foreign exchange (FX) markets, a *symmetry hint* has been used to regularize models against noisy financial data, leading to significantly improved annualized returns. Generative Hints is explicitly different from previous iterations of hints in its formulation of using a generative model to represent the input space to learn the functional properties.

3 WHAT ARE HINTS?

3.1 PROBLEM STATEMENT

We begin by defining the standard supervised learning setup. Let$f : X \to Y$ denote the true underlying function, where X is the input space and Y is the output space. Let D_{train} and D_{test} denote the training and test sets, respectively. In the case of image classification, X corresponds to the space of images and Y to the space of class probability distributions.

Intuitively, a *hint* encodes domain knowledge about a target function f by specifying how the output should behave under certain input transformations. For instance, in image classification, we know that flipping an image horizontally should not change its class label distribution. Rather than learning such properties implicitly form labeled data, hints allows us to enforce them explicitly during training. We now formalize this notion, beginning with the general concept of hints and the specializing with invariance hints.

Definition 1 (Invariance Hint)**.** A *hint* is a constraint on the target function f expressed through input transformations. Formally, let $h : X \to X$ be a transformation function and $R : Y \times Y \to \{0, 1\}$ be a binary relation on the output space. A hint specifies that for any input $x \in X$ the relationship $R(f(x), f(h(x))) = 1$ must hold. An *invariance hint* is a hint where the relationship R enforces output equality, specifying that for any $x \in X$,

$$f(h(x)) = f(x).$$

While data augmentation and our hint-based approach both leverage invariance, they differ fundamentally in their mechanism. Data augmentation applies transformations to *labeled* training examples $(x, y) \in D_{train}$, creating additional training pairs $(h(x), y)$ and rely on the supervised loss to implicitly teach invariance. In contrast, our method applies hints to *unlabeled* virtual examples (defined in Section 3.2) and uses an explicit auxiliary loss to enforce the invariance property directly. In practice, we apply both techniques simultaneously, with data augmentation operating on the labeled training data and hints providing additional invariance signal through virtual examples.

3.2 Enforcing Hints through Virtual Examples

We enforce invariance hints on *virtual examples*, which are unlabeled synthetic inputs from a generative model trained on the input distribution. Virtual examples were originally introduce in (Abu-Mostafa, 1995) for tabular data, where they were sampled from Gaussian noise. We adapt this concept to high-dimensional vision tasks by sampling from a learned generative model, ensuring the virtual examples are realistic and representative of the input image distribution.

Definition 2 (Invariance Hint on Virtual Examples)**.** Given a generative model G_θ and transformation $h : X \to X$, an *invariance hint on virtual examples* enforces that for any $z \sim p(z)$,

$$f(h(G_\theta(z))) = f(G_\theta(z)).$$

We instantiate our framework with three invariance hints commonly leveraged in data augmentation, corresponding to properties that image classification naturally respect.

Definition 3 (Spatial Invariance Hint)**.** Let $h_{\text{sp}} : X \to X$ denote a spatial transformation that applies a random horizontal flip with probability p_{flip}, translation within $\pm t$ pixels, and rotation within $\pm r$ degrees. A *spatial invariance hint* enforces that for any virtual example x_v sampled from G_θ,

$$f(h_{\text{sp}}(x_v)) = f(x_v).$$

Definition 4 (Photometric Invariance Hint)**.** Let $h_{\text{ph}} : X \to X$ denote a photometric transformation that adjusts brightness, contrast, and saturation by a multiplicative factor sampled uniformly from $[1-\alpha, 1+\alpha]$, where $\alpha > 0$ controls the strength of the transformation. A *photometric invariance hint* enforces that for any virtual example x_v sampled from G_θ,

$$f(h_{\text{ph}}(x_v)) = f(x_v).$$

Definition 5 (Cropping Invariance Hint)**.** Let $h_{\text{cr}} : X \to X$ denote a random crop operator that extracts a subregion of size $b \times b$ from an $a \times a$ image, where $b < a$. A *cropping invariance hint* enforces that for any virtual example x_v sampled from G_θ,

$$f(h_{\text{cr}}(x_v)) = f(x_v).$$

3.3 Training Generative Models Efficiently

We use StyleGAN3 from Karras et al. (2021) as our generative model to produce virtual examples. This model generates unlabeled images from the input distribution without class conditioning. While other generative models (ex. diffusion models) could be used, StyleGAN3 provided a favorable balance between sampling efficiency and image quality.

Training generative models in limited data settings requires careful data-efficient strategies to prevent overfitting. We leverage adaptive discriminator augmentation (ADA) from Karras et al. (2020a), which adjusts the strength of data augmentations dynamically based on overfitting signals, improving image quality in low-data regimes.We extend ADA with a curriculum learning approach where the full details can be seen in the Appendix.

3.4 Hint Loss Function

To enforce invariance, we measure prediction similarity using symmetric KL divergence with temperature scaling. For virtual example x_v with predictions $p = \hat{f}(x_v)$ and $q = \hat{f}(h(x_v))$, the temperature-scaled distributions are $\tilde{p} = \text{softmax}(p/T)$ and $\tilde{q} = \text{softmax}(q/T)$. The hint loss is:

$$\mathcal{L}_{\text{hint-ce}}(\tilde{p}, \tilde{q}) = \frac{1}{2}\Big(\text{KL}(\tilde{p} \,\|\, \tilde{q}) + \text{KL}(\tilde{q} \,\|\, \tilde{p})\Big).$$

For MSE-based training, we use: $\mathcal{L}_{\text{hint-mse}}(y_v, y'_v) = \frac{1}{d}\sum_{i=1}^{d}(y_{v,i} - y'_{v,i})^2$.

3.5 Training Algorithm

Our approach optimizes both classification loss on labeled data from D_{train} and hint loss on unlabeled virtual examples from G_θ, alternating between objectives each batch (Algorithm 1). The virtual examples are generated on-the-fly, ensuring diversity.

Algorithm 1 Generative Hints Training Algorithm

Input: Training set $\mathcal{D}_{\text{train}} = \{(x_i, y_i)\}_{i=1}^{N}$
Classifier $\hat{f}$, hint transformation $h : X \to X$, generator G_θ
Losses $\mathcal{L}_{\text{class}}$, $\mathcal{L}_{\text{hint}}$, coefficient α
Number of epochs E
for epoch $e = 1$ **to** E **do**
  for mini-batch $B \subset \mathcal{D}_{\text{train}}$ **do**
    Update $\hat{f}$ on B using $\mathcal{L}_{\text{class}}$ applying transformation h
    Sample $z \sim p(z)$ and set $x_v \sim G_\theta(z)$
    Compute $x'_v y_v \leftarrow \hat{f}(x_v)$ and $y'_v \leftarrow \hat{f}(x'_v)$
    Update $\hat{f}$ using $\alpha \mathcal{L}_{\text{hint}}(y_v, y'_v)$
  end for
end for

Figure 1: Virtual examples for each dataset: Stanford Cars (top left), CUB-200-2011 (top right), FGVC Aircraft (bottom left), and Oxford Flowers (bottom right). Each panel shows (left to right) an original training image from D_{train}, a virtual example x_v sampled from the generative model G_θ, and the transformed image $h_s(x_v)$ after applying the spatial invariance hint.

We introduce a coefficient α to scale the hint loss, controlling its relative weight compared to the classification objective. This weighting is necessary because the gradients and learning dynamics of the two objectives can differ significantly. In our experiments, we used a fixed α which provides stable and consistent improvements across datasets and architectures.

4 Experiments and Results

We evaluated our method on four popular fine-grained visual classification datasets: Stanford Cars (Krause et al., 2013), CUB-200-2011 (Wah et al., 2011), FGVC Aircraft (Maji et al., 2013), and Oxford Flowers (Nilsback & Zisserman, 2008). Experiments were conducted using two model architectures: ViT-B (Dosovitskiy et al., 2021; Vaswani et al., 2017) and Swin-B (Liu et al., 2021), chosen for their strong performance on fine-grained classification and to demonstrate the generality of our approach across architectures. We evaluated three hint types: spatial invariance, photometric invariance, and croppign invariance.

We further evaluated generative hints in a medical imaging setting using the CheXpert dataset (Irvin et al., 2019) with a ResNet50 (He et al., 2016), employing mean squared error as the training objective. Finally, we performed ablation studies to examine the impact of generative model quality on classification performance. All experiments were conducted on a single NVIDIA H100 GPU, training both the generative and classification models.

4.1 GENERATIVE MODEL TRAINING

We trained StyleGAN3 at 512×512 resolution on each dataset using ADA (see Appendix for hyperparameters). We trained for 5000 kimgs and selected checkpoints with best FID: 5.29 (Stanford Cars), 4.73 (FGVC Aircraft), 7.04 (CUB-200-2011), 12.62 (Oxford Flowers). Figure 1 shows virtual examples from each dataset and Table 1 summarizes the statistics of each dataset and the quality of the trained generative models, measured using Fréchet Inception Distance (FID) (Heusel et al., 2017).

Table 1: Dataset statistics for the four fine-grained visual classification benchmarks. FID is measured for StyleGAN3 trained on each training set, used for virtual example generation.

Dataset	Classes	Training Size	FID
Stanford Cars	196	8,144	5.29
FGVC Aircraft	100	6,800	4.73
CUB-200-2011	200	5,994	7.04
Oxford Flowers	102	2,040	12.62

4.2 FINE-GRAIN VISION CLASSIFICATION TRAINING RESULTS

Experimental Design To isolate the contribution of generative hints, we carefully control for the effect of data augmentation. Our baseline applies the same transformations used in our hints (spatial, photometric, and cropping) as standard data augmentation during training on labeled data. The generative hints condition uses these same augmentations on the training data, while *additionally* enforcing invariance constraints on unlabeled virtual examples via the hint loss $\mathcal{L}_{\text{hint}}$. This ensures that any performance improvement is attributable to the explicit invariance enforcement on virtual examples rather than the transformations themselves. Both conditions are trained with identical hyperparameters, differing only in the addition of the hint loss on virtual examples.

Model and Training Setup We evaluated ViT-B (patch size 16) and Swin-B (patch size 4), both pretrained on ImageNet-1k. All experiments used 384×384 resolution, batch size 32, AdamW optimizer (lr=0.0001), and trained for 200 epochs with cosine annealing. Hyperparameters were optimized for baselines using data augmentation and kept fixed for hint training.

Augmentation and Hints Spatial: horizontal flip (50% baseline, 100% hints), translation ($\pm 5\%$), rotation (± 5). Photometric: brightness/contrast/saturation factors in $[0.8, 1.2]$. Cropping: resize to 448^2, crop to 384^2. We used $\alpha = 1.0, T = 0.8$ (ViT-B) and $\alpha = 50.0, T = 0.8$ (Swin-B), tuned on Stanford Cars using the photometric hint and fixed for other datasets and hints. Each experiment was carried out with 5 seeds.

For generative hints training, we sampled one virtual example per real training example in each batch from the StyleGAN3 generative model (Section 5.1) and enforced the corresponding hint property. We performed a sweep over the temperature $T \in \{0.5, 0.8, 1.0, 1.2, 1.5, 2.0\}$ and hint loss coefficient $\alpha \in \{1.0, 5.0, 10.0, 25.0, 50.0, 75.0, 100.0\}$, tuning only on the Stanford Cars dataset for the photometric hint. For the final experiments, we used $\alpha = 1.0$ and $T = 0.8$ for ViT-B, and $\alpha = 50.0$ and $T = 0.8$ for Swin-B; these values were then kept fixed for all other datasets. Tables 11 and 12 in Appendix A.4 present an ablation study showing the effect of tuning α and T independently. Each experiment was run with 5 random seeds.

Results. Table 2 reports top-1 accuracy averaged across the three hint types (spatial, photometric, and cropping), with per-hint results provided in the Appendix. We also report the hint loss computed on virtual examples using the symmetric KL divergence with temperature $T = 1$. The baseline hint loss is computed by evaluating the trained baseline model on virtual examples and their transformations, measuring the implicit invariance learned through augmentation alone.

We observe consistent improvements across all datasets and both architectures through the use of generative hints. The substantial reduction in hint loss on virtual examples (from $\sim$0.2-0.7 down to $\sim 10^{-7}$-10^{-4}) demonstrates that explicit invariance enforcement successfully teaches the model to respect the specified transformations. While data augmentation alone provides some implicit invariance (as reflected in the baseline hint loss), the explicit hint objective on virtual examples

Table 2: Top-1 accuracy (Acc.) and hint loss (Hint L.) on virtual examples for ViT-B and Swin-B models. Results averaged over spatial, photometric, and cropping hints, with mean across 5 seeds. Baseline hint loss measures implicit invariance from augmentation; hints explicitly enforce invariance on virtual examples.

Dataset	ViT-B		ViT-B w/ Hints		Swin-B		Swin-B w/ Hints	
	Acc.	Hint L.	Acc.	Hint L.	Acc.	Hint L.	Acc.	Hint L.
Stanford Cars	90.15	0.714	**90.77**	**4.4e-05**	92.30	0.749	**92.95**	**2.3e-07**
FGVC Aircrafts	83.77	0.722	**85.09**	**1.8e-07**	91.17	0.772	**91.53**	**1.9e-07**
CUB-200-2011	88.09	0.571	**88.75**	**4.4e-05**	90.30	0.460	**90.88**	**2.9e-07**
Oxford Flowers	98.99	0.196	**99.46**	**8.5e-04**	99.60	0.176	**99.67**	**3.8e-06**

Table 3: Classification MSE loss across pathologies on CheXpert, with and without generative hints. Mean across 5 seeds. *Percent Gain* represents the relative MSE reduction from baseline.

Pathology	Baseline	w/ Hints	% Gain
No Finding	**0.636**	0.639	-0.472%
Enlarged Cardiomediastinum	0.719	**0.704**	2.086%
Cardiomegaly	0.339	**0.337**	0.590%
Lung Opacity	0.795	**0.784**	1.384%
Pneumonia	0.797	**0.781**	2.008%
Pleural Effusion	**0.423**	0.425	-0.473%
Pleural Other	0.876	**0.864**	1.370%
Fracture	0.673	**0.660**	1.932%
Support Devices	0.983	**0.952**	3.154%
Average Gain			1.286 %

substantially improves both invariance alignment and classification performance. This validates our hypothesis that virtual examples provide additional learning signal beyond what is available through augmentation on labeled data alone.

4.3 Medical Imaging: CheXpert

To evaluate the robustness and generality of our approach, we extended our experiments to medical imaging using the CheXpert dataset (Irvin et al., 2019). CheXpert is a large-scale chest radiograph dataset collected from Stanford Hospital, containing 224,316 X-rays from 65,240 patients, annotated for 14 common thoracic pathologies. For our experiments, we used 9 categories: No Finding, Enlarged Cardiomediastinum, Cardiomegaly, Lung Opacity, Pneumonia, Pleural Effusion, Pleural Other, Fracture, and Support Devices. Labels are automatically extracted from radiology reports using a rule-based NLP system.

Generative Model We trained StyleGAN3 at 256×256 on the full dataset (batch size 16, lr_G=0.0025, lr_D=0.001, $\gamma = 8.0$), achieving FID 4.38 for 5000 kimgs (thousand images processed) and selected the checkpoint with the best FID.

Classification Setup We used ResNet50 with ImageNet pretraining, 256×256 grayscale inputs, MSE loss for both classification and hints, batch size 64, Adam optimizer (lr=10^{-5}), 5 epochs. Baseline used spatial augmentation (translation/rotation 0-5%); hints added spatial invariance enforcement ($\alpha = 0.1$) using the transformations on virtual examples. We used only spatial invariance for this dataset, as photometric hints are inappropriate for X-rays (brightness/contrast variations can be clinically meaningful) and crop hints risk removing diagnostic information. Each experiment was run for 5 random seeds.

Results Table 3 reports classification MSE loss for each pathology. Generative hints improve performance on 7 out of 9 pathologies, with an average MSE reduction of 1.286%. While two categories (No Finding, Pleural Effusion) show slight increases in MSE, the overall trend demonstrates the benefit of explicit invariance enforcement. The improvements are particularly notable for Support Devices (3.154%), Pneumonia (2.008%), and Enlarged Cardiomediastinum (2.086%). These results demonstrate that generative hints generalize across domains (natural images to medical imaging), architectures (transformers to CNNs), and objectives (cross-entropy to MSE regression).

4.4 Effect of Generator Quality on Hint Learning

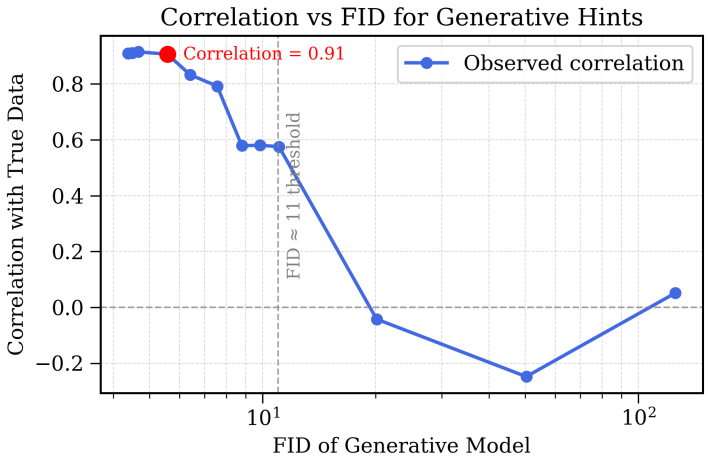

Figure 2: Correlation between generative hint loss on virtual examples and hint loss on real training data, plotted against FID. The horizontal dashed line marks zero correlation; the vertical dashed line marks the FID threshold ($\sim$11) where the generative model begins providing meaningful signal. The red point marks FID 5.58, where correlation reaches 0.91.

We conducted experiments to determine the quality of generative model required for effective hint learning. Specifically, we sought to identify the FID threshold at which the generative model sufficiently captures the input distribution such that hints learned on virtual examples transfer to real training data.

Experimental Setup. Following the CheXpert setup from Section 5.3, we trained classification models using generative hints with generators of varying quality (FID scores). Crucially, models were trained *only* on virtual examples using the hint loss (without any data augmentation on real training data), allowing us to isolate how well the generative model captures the true data distribution. We computed the correlation between hint loss on virtual examples and hint loss on real training data across 120 checkpoints sampled throughout 5 epochs of training. High correlation indicates that learning on virtual examples reflects learning on real data.

Results. Figure 2 shows correlation versus FID. At FID values above 50, correlation is near zero, indicating that low-quality generative models provide no meaningful learning signal. Once FID drops below 11, correlation becomes substantial, reaching 0.91 at FID 5.58. Beyond this point, further improvements in FID yield diminishing returns in correlation.

To understand how generator quality affects final classification performance, we ran an additional ablation using Swin-B with the photometric hint on Stanford Cars (Table 13). Performance remains stable and beneficial when FID is below $\sim$12, but degrades once FID exceeds 15. These results indicate that while high-quality generative models (FID $<$ 10) are ideal, moderate-quality models ($10 <$ FID < 15) still provide meaningful benefits for hint learning. The full table can be observed in Appendix A.5.

5 Conclusion

We proposed a method to reformulate supervised classification as semi-supervised learning by treating data synthesized from a generative model as unlabeled data, enabling models to learn functional properties through virtual examples. Our evaluations across fine-grained visual classification and medical imaging domains showed that generative hints consistently outperformed traditional data augmentation when learning the same properties, without requiring perfect generative models.

Future work includes developing dynamic schedulers to adapt objective weights, designing embedding hints over latent representations for properties difficult to encode via augmentation, and extending to object detection and segmentation tasks. This work establishes generative hints as a

versatile tool for domain knowledge injection, opening new avenues for explicit regularization in fully labeled settings.

REFERENCES

Yaser S. Abu-Mostafa. Learning from hints in neural networks. *Journal of Complexity*, 6(2):192–198, 1990. doi: 10.1016/0885-064X(90)90024-4.

Yaser S. Abu-Mostafa. Hints. *Neural Computation*, 7:639–671, July 1995.

Antreas Antoniou, Amos Storkey, and Harrison Edwards. Data augmentation generative adversarial networks. In *arXiv preprint arXiv:1711.04340*, 2017.

Shekoofeh Azizi, Simon Kornblith, Chitwan Saharia, Mohammad Norouzi, and David J Fleet. Synthetic data from diffusion models improves imagenet classification. *Transactions on Machine Learning Research*, 4:1–15, 2023. URL https://arxiv.org/abs/2304.08466.

Florian Bordes, Norman Mu, Boi Faltings, and et al. Synthetic data from diffusion models improves imagenet classification. In *arXiv preprint arXiv:2304.08466*, 2023.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 248–255. IEEE, 2009.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*, 2021.

Maayan Frid-Adar, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. In *arXiv preprint arXiv:1803.01229*, 2018.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, volume 30, pp. 6626–6637, 2017.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020.

Zixuan Huang, Yuchen Yang, Chen Liang, and et al. Ttida: Controllable generative data augmentation via text-to-text and text-to-image models. In *arXiv preprint arXiv:2304.08821*, 2023.

Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoo, Robyn Ball, Katie Shpanskaya, Jayne Seekins, David A. Mong, Safwan S. Halabi, Jesse K. Sandberg, Ricky Jones, David B. Larson, Curtis P. Langlotz, Bhavik N. Patel, Matthew P. Lungren, and Andrew Y. Ng. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 590–597, 2019. doi: 10.1609/aaai.v33i01.3301590. URL https://doi.org/10.1609/aaai.v33i01.3301590. https://stanfordmlgroup.github.io/competitions/chexpert/.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4401–4410, 2019.

Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. In *Proc. NeurIPS*, 2020a.

Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8110–8119, 2020b.

Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-free generative adversarial networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 852–863, 2021.

Diederik P Kingma, Danilo J Rezende, Shakir Mohamed, and Max Welling. Semi-supervised learning with deep generative models. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 3581–3589, 2014. URL https://papers.nips.cc/paper/5352-semi-supervised-learning-with-deep-generative-models.

Jonathan Krause, Jia Deng, Michael Stark, and Li Fei-Fei. Collecting a large-scale dataset of fine-grained cars. In *Proceedings of the Second Workshop on Fine-Grained Visual Categorization (FGVC2)*, 2013. URL https://ai.stanford.edu/~jkrause/papers/fgvc13.pdf. https://ai.stanford.edu/~jkrause/papers/fgvc13.pdf.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Zhuliang Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10012–10022, 2021.

Dwarikanath Mahapatra and Zongyuan Ge. Unified framework for histopathology image augmentation and classification via generative models. In *arXiv preprint arXiv:2212.09977*, 2022.

Subhransu Maji, Juho Kannala, Esa Rahtu, Matthew Blaschko, and Andrea Vedaldi. Fine-grained visual classification of aircraft. Technical Report 1306.5151, arXiv, 2013. URL https://arxiv.org/abs/1306.5151.

Maria-Elena Nilsback and Andrew Zisserman. Automated flower classification over a large number of classes. In *Proceedings of the 6th Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP)*, pp. 722–729. ACM, 2008. URL https://www.robots.ox.ac.uk/~vgg/data/flowers/102/.

Luis Perez and Jason Wang. The effectiveness of data augmentation in image classification using deep learning. *CoRR*, abs/1712.04621, 2017. URL http://arxiv.org/abs/1712.04621.

Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. In *International Conference on Learning Representations (ICLR)*, 2016. URL https://arxiv.org/abs/1511.06434.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bjorn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10684–10695, 2022.

Connor Shorten and Taghi M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):60, 2019. doi: 10.1186/s40537-019-0197-0.

Yang Song, Jia Meng, and Stefano Ermon. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations (ICLR)*, 2021. URL https://arxiv.org/abs/2011.13456.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 5998–6008. Curran Associates, Inc., 2017. URL https://arxiv.org/abs/1706.03762.

C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The caltech-ucsd birds-200-2011 dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011. URL http://www.vision.caltech.edu/visipedia/CUB-200-2011.html.

Jianhao Yuan, Jie Zhang, Shuyang Sun, Philip Torr, and Bo Zhao. Real-fake: Effective training data synthesis through distribution matching. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2024. URL https://openreview.net/forum?id=svIdLLZpsA.

Kai Zhang, Ming Wang, Jian Liu, and et al. Data augmentation for image classification using generative ai. In *arXiv preprint arXiv:2409.00547*, 2024.

Hang Zhao, Bowen Li, Tao Xu, and et al. Data augmentation using learned transformations for one-shot medical image segmentation. In *arXiv preprint arXiv:1902.09383*, 2019.

A Appendix

A.1 Datasets

We ran on the 4 datasets Stanford Cars, FGVC Aircrafts, CUB-200-2011, and Oxford Flowers. The datasets are all fine grain visual classification datasets with 100 or more classes. Full dataset specifications can be seen in Table 2. For datasets with a training/val/test split we combined the training and validation set, which is considered the standard for the datasets.

Table 4: Summary statistics for fine-grained visual classification datasets: number of classes, total image count, and standard train/test splits.

Dataset	# Classes	Total Images	Train Images	Test Images
Stanford Cars	196	16,185	8,144	8,041
FGVC Aircrafts	100	10,200	6,800	3,400
CUB-200-2011	200	11,788	5,994	5,794
Oxford Flowers 102	102	8,189	2,040	6,149

A.2 Generative Model Training

Generative models were trained according to the specifications listed in Table 5, using only the training split of each dataset. All models were trained until convergence with Adaptive Discriminator Augmentation (ADA) Karras et al. (2020a), in which augmentations are applied to images before being passed to the discriminator. Training was performed on a single NVIDIA H100 GPU and continued until Fréchet Inception Distance (FID) convergence Heusel et al. (2017). The augmentations used included: `xflip`, `rotate90`, `xint`, `scale`, `rotate`, `anisco`, `xfrac`, `brightness`, `contrast`, `lumaflip`, `hue`, and `saturation`. Full augmentation specifications are available in the official StyleGAN3 repository. The resulting generative model FIDs can be observed in Table 4.

Table 5: Training hyperparameters for the generative model (StyleGAN3 with ADA).

Hyperparameter	**Value**
Model Type	StyleGAN3
Resolution	512×512
Adaptive Discriminator Augmentation	Enabled
Mirror	Enabled
Optimizer	AdamW
Generator Learning Rate	0.0025
Discriminator Learning Rate	0.001
Batch Size	16
Gamma	4.0
Stop Condition	FID convergence

Table 6: The resulting StyleGAN3 models trained on each of the datasets including the FID achieved.

Dataset	# Classes	Train Images	FID
Stanford Cars	196	8,144	4.27
FGVC Aircrafts	100	6,800	4.72
CUB-200-2011	200	5,994	7.37
Oxford Flowers 102	102	2,040	12.62

A.3 Training Hyperparameters and Implementation Details

Both the ViT-B/16 and Swin-B transformer models were pretrained on ImageNet Deng et al. (2009). Table 7 reports the training hyperparameters, which were tuned to maximize baseline performance (without hints) and then kept fixed when applying generative hints. The only parameters modified for hint-based training were the hint loss weight α and the temperature T used in the symmetric KL divergence loss.

Data Augmentation. For baseline training (without hints), we applied the following augmentations: horizontal flip with probability $p = 0.5$, random rotation uniformly sampled from $[0°, 5°]$, and random translation uniformly sampled from $[0\%, 5\%]$ of image dimensions. When applying generative hints, we used identical augmentations on the training data, except that horizontal flip probability was increased to $p = 1.0$ for the spatial invariance hint to enforce complete flip invariance.

Hint-Specific Transformations. The three hint types used the following transformations:

- **Spatial Invariance:** Horizontal flip ($p = 1.0$), rotation $[0°, 5°]$, translation $[0\%, 5\%]$
- **Photometric Invariance:** Brightness, contrast, and saturation adjusted by factors uniformly sampled from $[0.8, 1.2]$ (i.e., $\pm 20\%$)
- **Cropping Invariance:** Images resized to 448×448 and randomly cropped to 384×384

Virtual Example Generation. During hint-based training, we generated one virtual example for each real sample per training batch by sampling from the pretrained StyleGAN3 model. Images were generated at 512×512 resolution and resized to 384×384 to match the model input size. We alternated between optimizing the supervised cross-entropy loss on real training data and the hint loss on virtual examples at every batch.

Hint Loss Configuration. We used symmetric KL divergence with temperature $T = 0.8$ for all experiments. Table 7 reports the best-performing α for each architecture, which was then used consistently across all datasets and hint types for that architecture. All training was performed on a single NVIDIA H100 GPU.

Table 7: Training and model hyperparameters for ViT-B/16 and Swin-B.

Hyperparameter	**ViT-B/16**	**Swin-B**
Resolution	384×384	384×384
Optimizer	AdamW	AdamW
Learning Rate	$1e{-4}$	$1e{-4}$
Weight Decay	0.01	0.01
Batch Size	32	32
Scheduler	Cosine Annealing	Cosine Annealing
Number of Epochs	200	200
Hint Loss Weight	1.0	50.0

Table 8: Top-1 accuracy using the photometric invariance hint for Stanford Cars, FGVC Aircraft, CUB-200-2011, and Oxford Flowers. Mean across 5 seeds. Bold indicates best performance. Average improvement: 0.76% (max: 2.10%).

Dataset	ViT-B Baseline	ViT-B w/ Hints	Swin-B Baseline	Swin-B w/ Hints
Stanford Cars	89.15	**90.29**	91.11	**92.13**
FGVC Aircrafts	82.51	**84.61**	90.23	**90.66**
CUB-200-2011	87.92	**88.37**	90.06	**90.58**
Oxford Flowers	99.14	**99.45**	99.58	**99.66**

Across all datasets and architectures, we observe:

1. **Photometric hints provide the largest gains** (avg. 0.76%), particularly on FGVC Aircraft where color/lighting variations are less semantically meaningful than spatial structure.

Table 9: Top-1 accuracy using the cropping invariance hint for Stanford Cars, FGVC Aircraft, CUB-200-2011, and Oxford Flowers. Mean across 5 seeds. Bold indicates best performance. Average improvement: 0.46% (max: 1.20%). Top-1 accuracy using the Crop hint for the Stanford Cars, FGVC Aircraft, CUB-200-2011, and Oxford Flowers datasets. Bold indicates the best performance for each dataset and model. Hints results in an average improvement of 0.38% (up to 1.2%).

Dataset	ViT-B Baseline	ViT-B w/ Hints	Swin-B Baseline	Swin-B w/ Hints
Stanford Cars	90.39	**90.43**	92.87	**93.18**
FGVC Aircrafts	82.37	**82.44**	90.74	**91.11**
CUB-200-2011	87.91	**89.11**	90.57	**90.96**
Oxford Flowers	98.89	**99.50**	99.62	**99.67**

Table 10: Top-1 accuracy using the spatial invariance hint for Stanford Cars, FGVC Aircraft, CUB-200-2011, and Oxford Flowers. Mean across 5 seeds. Bold indicates best performance. Average improvement: 0.63% (max: 1.78%).

Dataset	ViT-B Baseline	ViT-B w/ Hints	Swin-B Baseline	Swin-B w/ Hints
Stanford Cars	90.90	**91.58**	92.92	**93.53**
FGVC Aircrafts	86.43	**88.21**	92.55	**92.83**
CUB-200-2011	88.45	**88.76**	90.28	**91.11**
Oxford Flowers	98.94	**99.43**	99.61	**99.68**

2. **Spatial hints provide consistent improvements** (avg. 0.63%) across all datasets, demonstrating that explicit enforcement of flip and rotation invariance provides signal beyond standard augmentation.
3. **Cropping hints show more modest gains** (avg. 0.38%), likely because cropping can remove discriminative details in fine-grained classification, making strict invariance less appropriate.
4. **All three hints improve over baseline**, indicating that virtual examples successfully provide additional invariance signal regardless of the specific transformation type.
5. **Improvements are consistent across architectures**, with both ViT-B and Swin-B benefiting from hints, demonstrating generalizability.

The variability in per-hint performance suggests that hint selection could be optimized per-dataset based on domain knowledge about which invariances are most appropriate.

A.4 Alpha and Temperature Sensitivity Analysis

To understand the robustness of generative hints to hyperparameter choices, we conducted ablation studies on the two key parameters of our method: the hint loss weight α and the temperature parameter T used in the symmetric KL divergence loss. Both experiments were conducted using Swin-B on the Stanford Cars dataset with the photometric invariance hint, keeping all other hyperparameters fixed as specified in Table 7.

A.4.1 Effect of Hint Loss Weight α

The hint loss weight α controls the relative importance of the invariance objective compared to the supervised classification objective. Table 11 shows classification accuracy as a function of α. We observe that hints provide consistent improvements over the baseline (91.11%) across a wide range of α values from 1.0 to 100.0, with peak performance at $\alpha = 50.0$ (92.13%).

A.4.2 Effect of Temperature Parameter T

The temperature parameter T controls the sharpness of the probability distributions in the symmetric KL divergence loss. Lower temperatures produce sharper distributions, while higher temperatures produce smoother distributions. Table 12 shows classification accuracy as a function of T.

Table 11: Classification accuracy as a function of hint loss weight α for Swin-B on Stanford Cars with photometric hint. Baseline corresponds to training without hints. The method is robust across two orders of magnitude.

Setting	α	Accuracy
Baseline (No Hints)	–	91.11
w/Hints	1.0	91.52
w/Hints	5.0	91.77
w/Hints	10.0	91.85
w/Hints	25.0	91.63
w/Hints	50.0	92.13
w/Hints	75.0	91.84
w/Hints	100.0	92.02

Table 12: Classification accuracy as a function of temperature T for Swin-B on Stanford Cars with photometric hint. Baseline corresponds to training without hints. Moderate temperatures ($T \in [0.5, 1.2]$) provide stable performance.

Setting	T	Accuracy
Baseline (No Hints)	–	91.11
w/Hints	0.5	92.05
w/Hints	0.8	92.13
w/Hints	1.0	91.80
w/Hints	1.2	91.90
w/Hints	1.5	91.58
w/Hints	2.0	91.51

A.5 Generator Quality Ablation

To investigate how the quality of the generator affects downstream classification, we vary the generator's FID and measure its impact on accuracy using photometric hints with Swin-B on Stanford Cars. Table 13 shows that hints remain beneficial at moderate FID levels, with performance improving as generator quality increases, before plateauing or declining at lower quality levels.

Table 13: Effect of generator quality (FID) on classification accuracy using photometric hint with Swin-B on Stanford Cars. Mean across 5 seeds. Performance remains beneficial at moderate FID levels before declining as quality degrades.

Setting	FID ↓	Accuracy (%) ↑
Data Aug. (No Hints)	–	91.11
w/ Hints	30.83	91.09
w/ Hints	19.15	91.33
w/ Hints	14.49	91.59
w/ Hints	11.68	91.83
w/ Hints	5.29	92.13