1. ShahRukh Athar, Evgeny Burnaev, and Victor Lempitsky. Latent convolutional models. In ICLR, 2019

A new latent model of natural images that can be learned on large-scale datasets. The learning process provides a latent embedding for every image in the training dataset, as well as a deep convolutional network that maps the latent space to the image space. After training, the new model provides a strong and universal image prior for a variety of image restoration tasks such as large-hole inpainting, superresolution, and colorization. To model high-resolution natural images, our approach uses latent spaces of very high dimensionality (one to two orders of magnitude higher than previous latent image models). To tackle this high dimensionality, we use latent spaces with a special manifold structure (convolutional manifolds) parameterized by a ConvNet of a certain architecture. In the experiments, we compare the learned latent models with latent models learned by autoencoders, advanced variants of generative adversarial networks, and a strong baseline system using simpler parameterization of the latent space. Our model outperforms the competing approaches over a range of restoration tasks.

1. David Bau, Hendrik Strobelt, William Peebles, Jonas Wulff, Bolei Zhou, Jun-Yan Zhu, and Antonio Torralba. Semantic photo manipulation with a generative image prior. In SIGGRAPH, 2019

Despite the recent success of GANs in synthesizing images conditioned on inputs such as a user sketch, text, or semantic labels, manipulating the high-level attributes of an existing natural photograph with GANs is challenging for two reasons. First, it is hard for GANs to precisely reproduce an input image. Second, after manipulation, the newly synthesized pixels often do not fit the original image. In this paper, we address these issues by adapting the image prior learned by GANs to image statistics of an individual image. Our method can accurately reconstruct the input image and synthesize new content, consistent with the appearance of the input image. We demonstrate our interactive system on several semantic image editing tasks, including synthesizing new objects consistent with background, removing unwanted objects, and changing the appearance of an object. Quantitative and qualitative comparisons against several existing methods demonstrate the effectiveness of our method.

1. David Bau, Jun-Yan Zhu, Hendrik Strobelt, Bolei Zhou, Joshua B. Tenenbaum, William T. Freeman, and Antonio Torralba. Gan dissection: Visualizing and understanding generative adversarial networks. In ICLR, 2019

Generative Adversarial Networks (GANs) have achieved impressive results for many real-world applications. As an active research topic, many GAN variants have emerged with improvements in sample quality and training stability. However, visualization and understanding of GANs is largely missing. How does a GAN represent our visual world internally? What causes the artifacts in GAN results? How do architectural choices affect GAN learning? Answering such questions could enable us to develop new insights and better models. In this work, we present an analytic framework to visualize and understand GANs at the unit-, object-, and scene-level. We first identify a group of interpretable units that are closely related to concepts with a segmentation-based network dissection method. We quantify the causal effect of interpretable units by measuring the ability of interventions to control objects in the output. Finally, we examine the contextual relationship between these units and their surrounding by inserting the discovered concepts into new images. We show several practical applications enabled by our framework, from comparing internal representations across different layers, models, and datasets, to improving GANs by locating and removing artifact-causing units, to interactively manipulating objects in the scene. We will open source our interactive tools to help researchers and practitioners better understand their models.

1. Paul Hand and Vladislav Voroninski. Global guarantees for enforcing deep generative priors by empirical risk. IEEE Transactions on Information Theory, 2019

We examine the theoretical properties of enforcing priors provided by generative deep neural networks via empirical risk minimization. In particular we consider two models, one in which the task is to invert a generative neural network given access to its last layer and another in which the task is to invert a generative neural network given only compressive linear observations of its last layer. We establish that in both cases, in suitable regimes of network layer sizes and a randomness assumption on the network weights, that the non-convex objective function given by empirical risk minimization does not have any spurious stationary points. That is, we establish that with high probability, at any point away from small neighborhoods around two scalar multiples of the desired solution, there is a descent direction. Hence, there are no local minima, saddle points, or other stationary points outside these neighborhoods. These results constitute the first theoretical guarantees which establish the favorable global geometry of these non-convex optimization problems, and they bridge the gap between the empirical success of enforcing deep generative priors and a rigorous understanding of non-linear inverse problems.

1. Guang-Yuan Hao, Hong-Xing Yu, and Wei-Shi Zheng. Mixgan: learning concepts from different domains for mixture generation. In IJCAI, 2018.

absorbing different image concepts (e.g., content and style) from different domains and thus generating a new domain with learned concepts. In particular, we propose a mixture generative adversarial network (MIXGAN). MIXGAN learns concepts of content and style from two domains respectively, and thus can join them for mixture generation in a new domain, i.e., generating images with content from one domain and style from another. MIXGAN overcomes the limitation of current GAN-based models which either generate new images in the same domain as they observed in training stage, or require off-the-shelf content templates for transferring or translation. Extensive experimental results demonstrate the effectiveness of MIXGAN as compared to related state-of-the-art GAN-based models.

1. Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In CVPR, 2017

We investigate conditional adversarial networks as a general-purpose solution to image-to-image translation problems. These networks not only learn the mapping from input image to output image, but also learn a loss function to train this mapping. This makes it possible to apply the same generic approach to problems that traditionally would require very different loss formulations. We demonstrate that this approach is effective at synthesizing photos from label maps, reconstructing objects from edge maps, and colorizing images, among other tasks. Moreover, since the release of the pix2pix software associated with this paper, hundreds of twitter users have posted their own artistic experiments using our system. As a community, we no longer hand-engineer our mapping functions, and this work suggests we can achieve reasonable results without handengineering our loss functions either.

1. Ali Jahanian, Lucy Chai, and Phillip Isola. On the”steerability” of generative adversarial networks. In ICLR, 2020.

An open secret in contemporary machine learning is that many models work beautifully on standard benchmarks but fail to generalize outside the lab. This has been attributed to biased training data, which provide poor coverage over real world events. Generative models are no exception, but recent advances in generative adversarial networks (GANs) suggest otherwise - these models can now synthesize strikingly realistic and diverse images. Is generative modeling of photos a solved problem? We show that although current GANs can fit standard datasets very well, they still fall short of being comprehensive models of the visual manifold. In particular, we study their ability to fit simple transformations such as camera movements and color changes. We find that the models reflect the biases of the datasets on which they are trained (e.g., centered objects), but that they also exhibit some capacity for generalization: by "steering" in latent space, we can shift the distribution while still creating realistic images. We hypothesize that the degree of distributional shift is related to the breadth of the training data distribution. Thus, we conduct experiments to quantify the limits of GAN transformations and introduce techniques to mitigate the problem.

1. Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In ECCV, 2016

We consider image transformation problems, where an input image is transformed into an output image. Recent methods for such problems typically train feed-forward convolutional neural networks using a *per-pixel* loss between the output and ground-truth images. Parallel work has shown that high-quality images can be generated by defining and optimizing *perceptual* loss functions based on high-level features extracted from pretrained networks. We combine the benefits of both approaches, and propose the use of perceptual loss functions for training feed-forward networks for image transformation tasks. We show results on image style transfer, where a feed-forward network is trained to solve the optimization problem proposed by Gatys *et al.* in real-time. Compared to the optimization-based method, our network gives similar qualitative results but is three orders of magnitude faster. We also experiment with single-image super-resolution, where replacing a per-pixel loss with a perceptual loss gives visually pleasing results

1. Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. In ICLR, 2019

Despite recent progress in generative image modeling, successfully generating high-resolution, diverse samples from complex datasets such as ImageNet remains an elusive goal. To this end, we train Generative Adversarial Networks at the largest scale yet attempted, and study the instabilities specific to such scale. We find that applying orthogonal regularization to the generator renders it amenable to a simple "truncation trick," allowing fine control over the trade-off between sample fidelity and variety by reducing the variance of the Generator's input. Our modifications lead to models which set the new state of the art in class-conditional image synthesis. When trained on ImageNet at 128x128 resolution, our models (BigGANs) achieve an Inception Score (IS) of 166.5 and Frechet Inception Distance (FID) of 7.4, improving over the previous best IS of 52.52 and FID of 18.6.

1. Jingwen Chen, Jiawei Chen, Hongyang Chao, and Ming Yang. Image blind denoising with generative adversarial network based noise modeling. In CVPR, 2018

we consider a typical image blind denoising problem, which is to remove unknown noise from noisy images. As we all know, discriminative learning based methods, such as DnCNN, can achieve state-of-the-art denoising results, but they are not applicable to this problem due to the lack of paired training data. To tackle the barrier, we propose a novel two-step framework. First, a Generative Adversarial Network (GAN) is trained to estimate the noise distribution over the input noisy images and to generate noise samples. Second, the noise patches sampled from the first step are utilized to construct a paired training dataset, which is used, in turn, to train a deep Convolutional Neural Network (CNN) for denoising. Extensive experiments have been done to demonstrate the superiority of our approach in image blind denoising.

1. Xinyuan Chen, Chang Xu, Xiaokang Yang, Li Song, and Dacheng Tao. Gated-gan: Adversarial gated networks for multi-collection style transfer. TIP, 2018.

Style transfer describes the rendering of an image's semantic content as different artistic styles. Recently, generative adversarial networks (GANs) have emerged as an effective approach in style transfer by adversarially training the generator to synthesize convincing counterfeits. However, traditional GAN suffers from the mode collapse issue, resulting in unstable training and making style transfer quality difficult to guarantee. In addition, the GAN generator is only compatible with one style, so a series of GANs must be trained to provide users with choices to transfer more than one kind of style. In this paper, we focus on tackling these challenges and limitations to improve style transfer. We propose adversarial gated networks (Gated-GAN) to transfer multiple styles in a single model. The generative networks have three modules: an encoder, a gated transformer, and a decoder. Different styles can be achieved by passing input images through different branches of the gated transformer. To stabilize training, the encoder and decoder are combined as an auto-encoder to reconstruct the input images. The discriminative networks are used to distinguish whether the input image is a stylized or genuine image. An auxiliary classifier is used to recognize the style categories of transferred images, thereby helping the generative networks generate images in multiple styles. In addition, Gated-GAN makes it possible to explore a new style by investigating styles learned from artists or genres. Our extensive experiments demonstrate the stability and effectiveness of the proposed model for multi-style transfer.

1. Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In CVPR, 2018

Recent studies have shown remarkable success in image-to-image translation for two domains. However, existing approaches have limited scalability and robustness in handling more than two domains, since different models should be built independently for every pair of image domains. To address this limitation, we propose StarGAN, a novel and scalable approach that can perform image-to-image translations for multiple domains using only a single model. Such a unified model architecture of StarGAN allows simultaneous training of multiple datasets with different domains within a single network. This leads to StarGAN's superior quality of translated images compared to existing models as well as the novel capability of flexibly translating an input image to any desired target domain. We empirically demonstrate the effectiveness of our approach on a facial attribute transfer and a facial expression synthesis tasks.

1. Antonia Creswell and Anil Anthony Bharath. Inverting the generator of a generative adversarial network. TNNLS, 2018

Generative adversarial networks (GANs) learn a deep generative model that is able to synthesize novel, high-dimensional data samples. New data samples are synthesized by passing latent samples, drawn from a chosen prior distribution, through the generative model. Once trained, the latent space exhibits interesting properties that may be useful for downstream tasks such as classification or retrieval. Unfortunately, GANs do not offer an “inverse model,” a mapping from data space back to latent space, making it difficult to infer a latent representation for a given data sample. In this paper, we introduce a technique, inversion, to project data samples, specifically images, to the latent space using a pretrained GAN. Using our proposed inversion technique, we are able to identify which attributes of a data set a trained GAN is able to model and quantify GAN performance, based on a reconstruction loss. We demonstrate how our proposed inversion technique may be used to quantitatively compare the performance of various GAN models trained on three image data sets. We provide codes for all of our experiments in the website (<https://github.com/ToniCreswell/InvertingGAN>).

1. Zachary C Lipton and Subarna Tripathi. Precise recovery of latent vectors from generative adversarial networks. In ICLR Workshop, 2017. 2

Generative adversarial networks (GANs) transform latent vectors into visually plausible images. It is generally thought that the original GAN formulation gives no out-of-the-box method to reverse the mapping, projecting images back into latent space. We introduce a simple, gradient-based technique called stochastic clipping. In experiments, for images generated by the GAN, we precisely recover their latent vector pre-images 100% of the time. Additional experiments demonstrate that this method is robust to noise. Finally, we show that even for unseen images, our method appears to recover unique encodings.

1. Xiaodan Liang, Hao Zhang, Liang Lin, and Eric Xing. Generative semantic manipulation with mask-contrasting gan. In ECCV, 2018. 2

Despite the promising results on paired/unpaired image-to-image translation achieved by Generative Adversarial Networks (GANs), prior works often only transfer the low-level information (eg color or texture changes), but fail to manipulate high-level semantic meanings (eg, geometric structure or content) of different object regions. On the other hand, while some researches can synthesize compelling real-world images given a class label or caption, they cannot condition on arbitrary shapes or structures, which largely limits their application scenarios and interpretive capability of model results. In this work, we focus on a more challenging semantic manipulation task, aiming at modifying the semantic meaning of an object while preserving its own characteristics (eg viewpoints and shapes), such as cow $ ightarrow $ sheep, motor $ ightarrow $ bicycle, cat $ ightarrow $ dog. To tackle such large semantic changes, we introduce a contrasting GAN (contrast-GAN) with a novel adversarial contrasting objective which is able to perform all types of semantic translations with one category-conditional generator. Instead of directly making the synthesized samples close to target data as previous GANs did, our adversarial contrasting objective optimizes over the distance comparisons between samples, that is, enforcing the manipulated data be semantically closer to the real data with target category than the input data. Equipped with the new contrasting objective, a novel mask-conditional contrast-GAN architecture is proposed to enable disentangle image background with object semantic changes. Extensive qualitative and quantitative experiments on several semantic manipulation tasks on ImageNet and MSCOCO dataset show considerable performance gain by our contrast-GAN over other conditional GANs.

1. Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, ´ Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photorealistic single image super-resolution using a generative adversarial network. In CVPR, 2017

Despite the breakthroughs in accuracy and speed of single image super-resolution using faster and deeper convolutional neural networks, one central problem remains largely unsolved: how do we recover the finer texture details when we super-resolve at large upscaling factors? The behavior of optimization-based super-resolution methods is principally driven by the choice of the objective function. Recent work has largely focused on minimizing the mean squared reconstruction error. The resulting estimates have high peak signal-to-noise ratios, but they are often lacking high-frequency details and are perceptually unsatisfying in the sense that they fail to match the fidelity expected at the higher resolution. In this paper, we present SRGAN, a generative adversarial network (GAN) for image super-resolution (SR). To our knowledge, it is the first framework capable of inferring photo-realistic natural images for 4x upscaling factors. To achieve this, we propose a perceptual loss function which consists of an adversarial loss and a content loss. The adversarial loss pushes our solution to the natural image manifold using a discriminator network that is trained to differentiate between the super-resolved images and original photo-realistic images. In addition, we use a content loss motivated by perceptual similarity instead of similarity in pixel space. Our deep residual network is able to recover photo-realistic textures from heavily downsampled images on public benchmarks. An extensive mean-opinion-score (MOS) test shows hugely significant gains in perceptual quality using SRGAN. The MOS scores obtained with SRGAN are closer to those of the original high-resolution images than to those obtained with any state-of-the-art method.

1. Guillaume Lample, Neil Zeghidour, Nicolas Usunier, Antoine Bordes, Ludovic Denoyer, and Marc’Aurelio Ranzato. Fader networks: Manipulating images by sliding attributes. In NeurIPS, 2017.

a new encoder-decoder architecture that is trained to reconstruct images by disentangling the salient information of the image and the values of attributes directly in the latent space. As a result, after training, our model can generate different realistic versions of an input image by varying the attribute values. By using continuous attribute values, we can choose how much a specific attribute is perceivable in the generated image. This property could allow for applications where users can modify an image using sliding knobs, like faders on a mixing console, to change the facial expression of a portrait, or to update the color of some objects. Compared to the state-of-the-art which mostly relies on training adversarial networks in pixel space by altering attribute values at train time, our approach results in much simpler training schemes and nicely scales to multiple attributes. We present evidence that our model can significantly change the perceived value of the attributes while preserving the naturalness of images.

1. Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In NeurIPS, 2018

Flow-based generative models are conceptually attractive due to tractability of the exact log-likelihood, tractability of exact latent-variable inference, and parallelizability of both training and synthesis. In this paper we propose Glow, a simple type of generative flow using invertible 1x1 convolution. Using our method we demonstrate a significant improvement in log-likelihood and qualitative sample quality. Perhaps most strikingly, we demonstrate that a generative model optimized towards the plain log-likelihood objective is capable of efficient synthesis of large and subjectively realistic-looking images.

1. Fangchang Ma, Ulas Ayaz, and Sertac Karaman. Invertibility of convolutional generative networks from partial measurements. In NeurIPS, 2018.

new theoretical results on convolutional generative neural networks, in particular their invertibility (ie, the recovery of input latent code given the network output). The study of network inversion problem is motivated by image inpainting and the mode collapse problem in training GAN. Network inversion is highly non-convex, and thus is typically computationally intractable and without optimality guarantees. However, we rigorously prove that, under some mild technical assumptions, the input of a two-layer convolutional generative network can be deduced from the network output efficiently using simple gradient descent. This new theoretical finding implies that the mapping from the low-dimensional latent space to the high-dimensional image space is bijective (ie, one-to-one). In addition, the same conclusion holds even when the network output is only partially observed (ie, with missing pixels). Our theorems hold for 2-layer convolutional generative network with ReLU as the activation function, but we demonstrate empirically that the same conclusion extends to multi-layer networks and networks with other activation functions, including the leaky ReLU, sigmoid and tanh.

1. Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a completely blind image quality analyzer. IEEE Signal Processing Letters, 2012

An important aim of research on the blind image quality assessment (IQA) problem is to devise perceptual models that can predict the quality of distorted images with as little prior knowledge of the images or their distortions as possible. Current state-of-the-art “general purpose” no reference (NR) IQA algorithms require knowledge about anticipated distortions in the form of training examples and corresponding human opinion scores. However we have recently derived a blind IQA model that only makes use of measurable deviations from statistical regularities observed in natural images, without training on human-rated distorted images, and, indeed without any exposure to distorted images. Thus, it is “completely blind.” The new IQA model, which we call the Natural Image Quality Evaluator (NIQE) is based on the construction of a “quality aware” collection of statistical features based on a simple and successful space domain natural scene statistic (NSS) model. These features are derived from a corpus of natural, undistorted images. Experimental results show that the new index delivers performance comparable to top performing NR IQA models that require training on large databases of human opinions of distorted images.

1. Yujun Shen, Ping Luo, Junjie Yan, Xiaogang Wang, and Xiaoou Tang. Faceid-gan: Learning a symmetry three-player gan for identity-preserving face synthesis. In CVPR, 2018.

Face synthesis has achieved advanced development by using generative adversarial networks (GANs). Existing methods typically formulate GAN as a two-player game, where a discriminator distinguishes face images from the real and synthesized domains, while a generator reduces its discriminativeness by synthesizing a face of photo-realistic quality. Their competition converges when the discriminator is unable to differentiate these two domains. Unlike two-player GANs, this work generates identity-preserving faces by proposing FaceID-GAN, which treats a classifier of face identity as the third player, competing with the generator by distinguishing the identities of the real and synthesized faces (see Fig. 1). A stationary point is reached when the generator produces faces that have high quality as well as preserve identity. Instead of simply modeling the identity classifier as an additional discriminator, FaceID-GAN is formulated by satisfying information symmetry, which ensures that the real and synthesized images are projected into the same feature space. In other words, the identity classifier is used to extract identity features from both input (real) and output (synthesized) face images of the generator, substantially alleviating training difficulty of GAN. Extensive experiments show that FaceID-GAN is able to generate faces of arbitrary viewpoint while preserve identity, outperforming recent advanced approaches.

1. Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.

we investigate the effect of the convolutional network depth on its accuracy in the large-scale image recognition setting. Our main contribution is a thorough evaluation of networks of increasing depth using an architecture with very small (3x3) convolution filters, which shows that a significant improvement on the prior-art configurations can be achieved by pushing the depth to 16-19 weight layers. These findings were the basis of our ImageNet Challenge 2014 submission, where our team secured the first and the second places in the localisation and classification tracks respectively. We also show that our representations generalise well to other datasets, where they achieve state-of-the-art results. We have made our two best-performing ConvNet models publicly available to facilitate further research on the use of deep visual representations in computer vision.

1. Patricia L Suarez, Angel D Sappa, and Boris X Vintimilla. ´ Infrared image colorization based on a triplet dcgan architecture. In CVPR Workshop, 2017

This paper proposes a novel approach for colorizing near infrared (NIR) images using a Deep Convolutional Generative Adversarial Network (GAN) architecture. The proposed approach is based on the usage of a triplet model for learning each color channel independently, in a more homogeneous way. It allows a fast convergence during the training, obtaining a greater similarity between the colored NIR image and the corresponding ground truth. The proposed approach has been evaluated with a large data set of NIR images and compared with a recent approach, which is also based on a GAN architecture where all the color channels are obtained at the same time.

1. Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In CVPR, 2018.

Deep convolutional networks have become a popular tool for image generation and restoration. Generally, their excellent performance is imputed to their ability to learn realistic image priors from a large number of example images. In this paper, we show that, on the contrary, the structure of a generator network is sufficient to capture a great deal of low-level image statistics prior to any learning. In order to do so, we show that a randomly-initialized neural network can be used as a handcrafted prior with excellent results in standard inverse problems such as denoising, superresolution, and inpainting. Furthermore, the same prior can be used to invert deep neural representations to diagnose them, and to restore images based on flash-no flash input pairs. Apart from its diverse applications, our approach highlights the inductive bias captured by standard generator network architectures. It also bridges the gap between two very popular families of image restoration methods: learning-based methods using deep convolutional networks and learning-free methods based on handcrafted image priors such as self-similarity.

1. Paul Upchurch, Jacob Gardner, Geoff Pleiss, Robert Pless, Noah Snavely, Kavita Bala, and Kilian Weinberger. Deep feature interpolation for image content changes. In CVPR, 2017

We propose Deep Feature Interpolation (DFI), a new data-driven baseline for automatic high-resolution image transformation. As the name suggests, DFI relies only on simple linear interpolation of deep convolutional features from pre-trained convnets. We show that despite its simplicity, DFI can perform high-level semantic transformations like" make older/younger"," make bespectacled"," add smile", among others, surprisingly well--sometimes even matching or outperforming the state-of-the-art. This is particularly unexpected as DFI requires no specialized network architecture or even any deep network to be trained for these tasks. DFI therefore can be used as a new baseline to evaluate more complex algorithms and provides a practical answer to the question of which image transformation tasks are still challenging after the advent of deep learning.