

InGAN: Capturing and Retargeting the “DNA” of a Natural Image

Assaf Shocher* Shai Bagon† Phillip Isola‡ Michal Irani*

*Dept. of Computer Science and Applied Math, The Weizmann Institute of Science

† Weizmann Artificial Intelligence Center (WAIC)

‡Computer Science and Artificial Intelligence Lab Massachusetts Institute of Technology

Project Website: <http://www.wisdom.weizmann.ac.il/~vision/ingan/>

Abstract

*Generative Adversarial Networks (GANs) typically learn a distribution of images in a large image dataset, and are then able to generate new images from this distribution. However, each natural image has its own internal statistics, captured by its unique distribution of patches. In this paper we propose an “Internal GAN” (InGAN) – an image-specific GAN – which trains on a single input image and learns its internal distribution of patches. It is then able to synthesize a plethora of new natural images of significantly different sizes, shapes and aspect-ratios – all with the same internal patch-distribution (same “DNA”) as the input image. In particular, despite large changes in global size/shape of the image, all elements inside the image maintain their local size/shape. InGAN is fully unsupervised, requiring no additional data other than the input image itself. Once trained on the input image, it can remap the input to any size or shape in a single feedforward pass, while preserving the same internal patch distribution. InGAN provides a unified framework for a variety of tasks, bridging the gap between textures and natural images.*¹

1. Introduction

Each natural image has its unique internal statistics: small patches (e.g., 5x5, 7x7) recur abundantly inside a single natural image [13, 36]. This patch recurrence was shown to form a *strong image-specific prior* for solving many ill-posed vision tasks in an unsupervised way [3, 6, 9, 8, 13, 29, 23, 2, 5]. In this paper we *capture and visualize* this unique image-specific patch-distribution, and map it to new target images of different sizes and shapes – all with the same internal patch distribution as the input image (which we loosely call “same DNA”).

For example, imagine you are given an input image, and you wish to transform it to a new image, of drastically different shape, size and aspect ratio. But you don’t want to distort any of its internal elements; you want to keep them all in their original size, shape, aspect ratio, and in the same relative position within the image. Such examples are shown in Fig. 1. Note that despite changing the global size and shape of the farmhouse image, the windows in the target images maintain their local size and shape. Rows of windows are automatically added/removed, and likewise for the number of windows in each row. Similarly, when the fruit-stand image in Fig. 1 is enlarged, more fruits are added in each fruit-box while keeping the size of each fruit the same; and vice versa – when the image grows smaller, the number of fruits grows smaller, while maintaining their size and their relative position within the image. Furthermore, note that the target image may not necessarily be rectangular.

How can this be done? One way to satisfy these criteria is to require that the distribution of patches in the target images match the distribution of patches in the input image, at multiple image scales. We propose Distribution-Matching as a new objective for “visual retargeting”. Note that we use the term retargeting here differently than its common use in image-retargeting methods [1, 4, 33]. Distribution-matching allows synthesizing new target images of different sizes and shapes – all with the same internal patch distribution as the input image.

A closely related work is the Bidirectional-Similarity of Simakov et al. [29]. The Bidirectional objective constrains the target image to contain only patches from the input image (“Visual Coherence”), and vice versa, the input should contain only patches from the target (“Visual Completeness”). Hence, no new artifacts are introduced in the target image and no critical information is lost either. Our new “Distribution Matching” formulation extends the Bidirectional-Similarity and goes beyond it in multiple ways: (i) It requires not only that all input patches be in the output (and vice versa), but also that the frequency of these patches remain the same. (ii) By matching distribu-

¹**Acknowledgement:** Funded by the European Research Council (ERC) under the Horizon 2020 research and innovation programme (grant No 788535). Also supported by a research grant from the Carolito Stiftung. Dr Bagon is a Robin Chemers Neustein Artificial Intelligence Fellow.

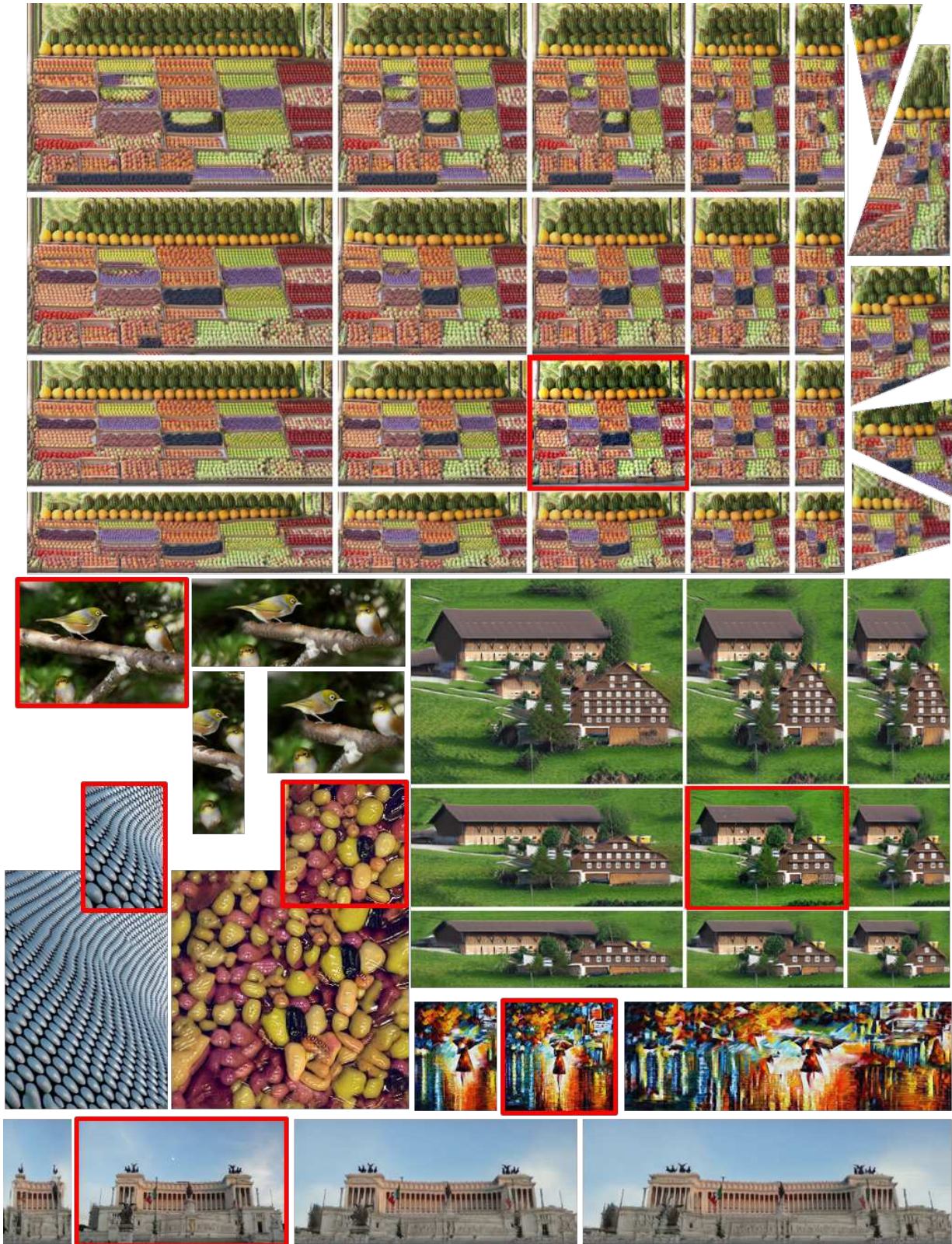


Figure 1: InGAN's Capabilities: (Top:) Once trained on an input image (marked by a *red frame*), InGAN can synthesize a plethora of new images of significantly different sizes/shapes/aspect-ratios all with the same “DNA” of the input image. All elements inside the image maintain their local size/shape and relative position. **Please view attached videos to see the continuum between different shapes/sizes/aspect-ratios.** (Bottom:) InGAN provides a unified treatment for a variety of different datatypes – single/multi-texture images, paintings, and complex natural images, all under a single umbrella.

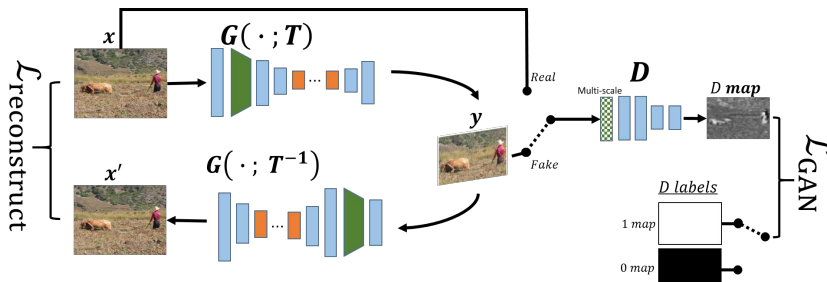


Figure 2: **InGAN Architecture:** InGAN consists of a Generator G that retargets input x to output y whose size/shape is determined by a geometric transformation T (top left). A multiscale discriminator D learns to discriminate the patch statistics of the fake output y from the true patch statistics of the input image (right). Additionally, we take advantage of G 's automorphism to reconstruct the input back from y using G and the inverse transformation T^{-1} (bottom left).

tions rather than individual patches, we can leverage recent advances in distribution modeling using Generative Adversarial Networks (GANs) [14]. (iii) A single forward pass through our trained network can generate target images of any size/shape/aspect ratio, without having to solve a new optimization problem for each desired target.

GANs can be understood as a tool for distribution matching [14]. A GAN typically learns a distribution of images in a large image dataset. It maps data sampled from one distribution to transformed data that is indistinguishable from a target distribution, $G : x \rightarrow y$ with $x \sim p_x$, and $G(x) \sim p_y$. We propose an “Internal GAN” (InGAN) – an **image-specific GAN** – which trains on a single input image and learns its unique internal distribution of patches. InGAN is fully unsupervised, requiring no training examples other than the input image. Unlike most GANs, which map between two different distributions, InGAN is an *automorphism*, $G : x \rightarrow x$, with p_x being the distribution of patches in the input image. Retargeting is achieved by modifying the size and shape of the output tensor, which changes the arrangement of patches, but not the distribution of patches.

Although this formulation is sufficient in theory to encourage both *Coherence* and *Completeness*, in practice we observe that completeness is often not achieved – many patches from the input image are omitted in the output (“mode collapse”). To ameliorate this, we introduce a second mechanism for encouraging completeness: it should be possible to reconstruct (“decode”) the input image from the output, i.e. $\|F(G(x)) - x\|$ should be small, where F is a second network trained to perform the reverse mapping. This objective encourages the mapping between input and retargeted output to be cycle-consistent [35], a desideratum that has recently come into widespread use and often improves the results of distribution matching problems. Since our proposed InGAN is an automorphism, we use G itself to perform the decoding, that is $\|G(G(x)) - x\|$ resulting in a novel **Encoder-Encoder** architecture.

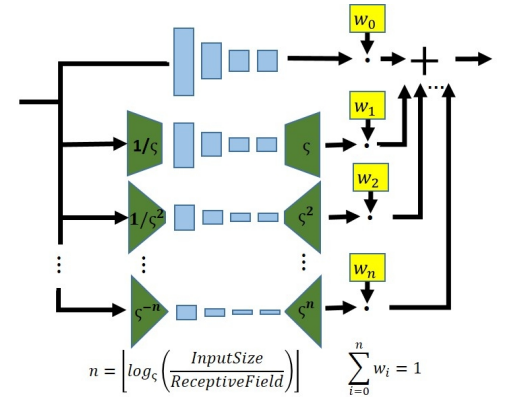


Figure 3: **Adaptive Multi-Scale Patch Discriminator**

Our results reinforce the recent finding that neural nets, when trained on a *single* image, can learn a useful representation of the internal statistics of that image. These representations can then be used to super-resolve the image [28], to inpaint patches removed from the image [31], or to synthesize textures from a sample texture image [18, 34, 30]. In particular, GANs trained on a single *texture* image were introduced by [18, 34]. A concurrent paper (at ICCV), SinGAN [26], also trains a GAN on a single natural image in order to achieve image manipulations. While they solve this problem using an unconditional GAN, which maps noise to images, our model is a conditional GAN, which maps images to images. Through InGAN, we further show that such image-specific internal statistics, encoded in a feedforward neural net, provides a single unified framework for a variety of new tasks/capabilities (Image-Retargeting, Image Summarization & Expansion, Texture-Synthesis, synthesizing Non-Rectangular outputs, etc.) Through its multi-scale discriminator, InGAN further provides a unified treatment for a variety of different datatypes (single/multi-texture images, paintings, and complex natural images), all under a single umbrella. While not guaranteed to provide state-of-the-art results compared to specialized methods optimized for a specific task/datatype, it compares favorably to them, and further gives rise to new applications.

Our contributions are several-fold:

- We define distribution-matching of patches as a criterion for visual retargeting and image manipulation.
- InGAN provides a unified-framework for various tasks and different datatypes, all with a single architecture.
- Once trained, InGAN can produce outputs of significantly different sizes, shapes, and aspect ratios, including non-rectangular output images.
- To the best of our knowledge, InGAN is the first to train a GAN on a single *natural* image (first appeared in arxiv 2018 [27]).
- The inherent symmetry of the challenge (an Automor-

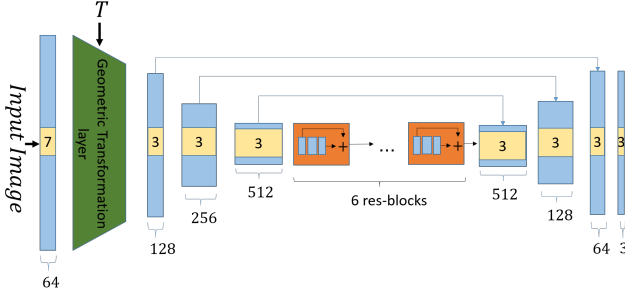


Figure 4: **Generator architecture:** G receives an input image x and a geometric transformation T which determines the size/shape of the output.

phism) gives rise to a new *Encoder-Encoder* architecture.

2. Method

2.1. Overview

Our InGAN is an image-conditional GAN (e.g., [17]) that maps an input image (as opposed to noise) to a remapped target output image. It uses a generator, G , a discriminator, D , and re-uses G for decoding/reconstructing the input given the output, as depicted in Fig. 2. Our formulation aims to achieve two properties: (i) **matching distributions:** The distribution of patches, across scales, in the synthesized image, should match that distribution in the original input image. This property is a generalization of both the *Coherence* and *Completeness* objectives of [29]. (ii) **localization:** The elements' locations in the generated image should generally match their relative locations in the original input image.

In detail, our method works as follows. Given an input image x , and a geometric transformation T (which determines the desired shape/size of the target output), G synthesizes a new image, $y = G(x; T)$. For example, T could be a scaling transformation, a skew, an affine transformation, or any other invertible geometric transformation. In our current implementation we allow T to be any desired homography (a 2D projective transformation). During training, T is randomly sampled at each iteration. Once trained, G can handle any desired transformation T (any homography).

The generator G trains to output an image y of size/shape/aspect-ratio specified by T that, at the patch level, is indistinguishable from the input image x , according to an adversarial discriminator D . We adopt the LSGAN [20] variant of this optimization problem: $G^* = \arg\min_G \max_D \mathcal{L}_{\text{GAN}}(G, D)$, where

$$\mathcal{L}_{\text{GAN}}(G, D) = \mathbb{E}_{x \sim \text{patches}(I)} [(D(x) - 1)^2 + D(G(x))^2]$$

The discriminator D and $\mathcal{L}_{\text{GAN}}$ encourage matching the patch distribution of $y = G(x; T)$ to that of x . D is fully convolutional: it outputs a map (rather than a scalar) where

each pixel depends only on its receptive field [7], thus it has all the patches of the original input x to train on. Using a multiscale D enforces patch distribution matching at each scale separately.

In practice using only $\mathcal{L}_{\text{GAN}}$ may result in mode collapse, i.e. the synthesized image consists of only a subset of patches of the original image (it is *coherent*) but many patches are missing (it is not *complete*). To ameliorate this mode collapse we take advantage of the automorphism of G and re-use G to reconstruct x back from the synthesized image y . The ℓ_1 reconstruction loss $\mathcal{L}_{\text{reconst}} = \|G(G(x; T); T^{-1}) - x\|_1$ encourages G to avoid mode collapse and maintain *completeness*. The overall loss function of InGAN is $\mathcal{L}_{\text{InGAN}} = \mathcal{L}_{\text{GAN}} + \lambda \cdot \mathcal{L}_{\text{reconst}}$.

Localization is implicitly encouraged through the choice of network architecture. The architecture is locally-connected rather than fully-connected (in particular, it is convolutional). This means that an output pixel at location $\{i, j\}$ can only depend on input pixels in a finite receptive field around that location in the input image. Nonlocal mappings, beyond a certain radius, are impossible with such an architecture. We also conjecture that simple local mappings are easier to learn than nonlocal mappings, and convolutional networks may naturally converge to these solutions [10].

2.2. Shape-flexible Generator

Fig. 4 shows the architecture of the generator G . The desired geometric transformation for the output shape T is treated as an additional input that is fed to G for every forward pass. A *parameter-free* transformation layer (green layer in Fig. 4) geometrically transforms the feature map to the desired output shape. Making the transformation layer parameter-free allows training G once to transform x to any size, shape or aspect ratio at test time.

The generator is fully-convolutional with an hourglass architecture and skip connections (U-net [24] architecture). The bottleneck consists of residual-blocks [15]. Downscaling is done by max pooling. Upscaling is done by nearest-neighbor resizing followed by a convolutional layer [22].

2.3. Multi-scale Patch Discriminator

We use a fully-convolutional patch discriminator D (Fig. 3), as introduced in [17]. The labels for the discriminator are maps (matrices of real/fake labels) of same size as the desired output y . Thus D grades each patch for how well it matches the patch distribution, rather than grading the entire synthesized image.

InGAN uses a multi-scale D (similar to [32]). This feature is significant: A single scale discriminator can only capture patch statistics of a specific size. Using a multiscale D matches the patch distribution over a range of patch sizes, capturing both fine-grained details as well as coarse struc-

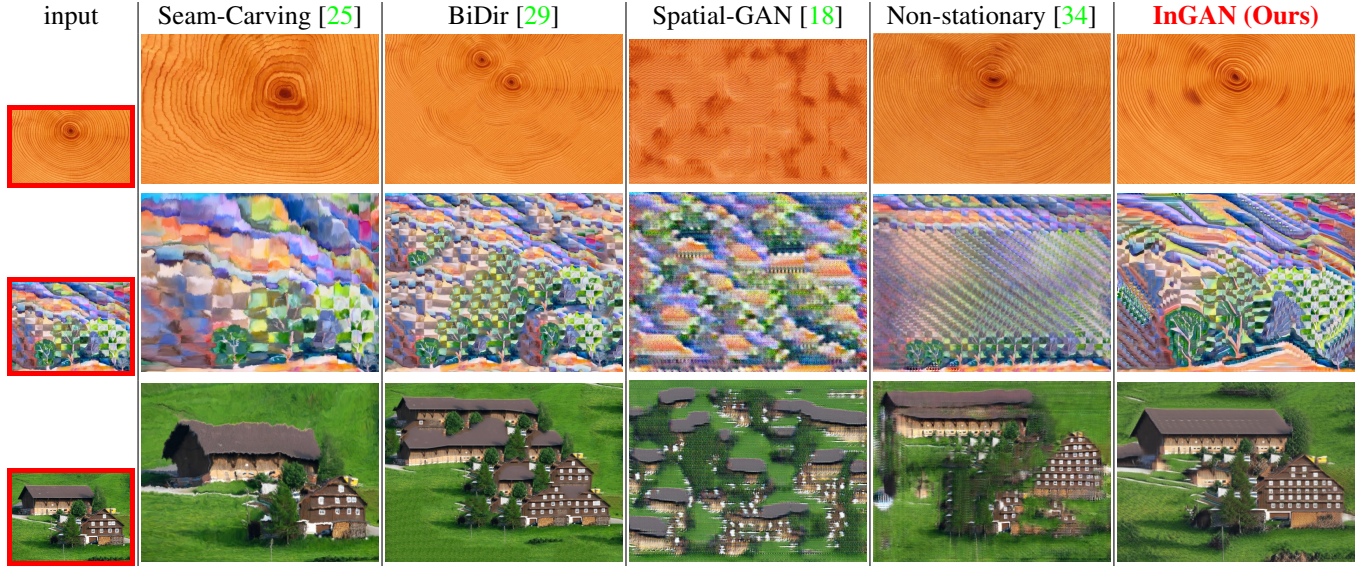


Figure 5: **Unified treatment for a range of datatypes** *InGAN* handles textures, paintings and natural images with a single architecture, whereas texture synthesis methods [18, 34] poorly handle natural images, and retargeting methods [25, 29] struggle with textures.

tures in the image. At each scale, the discriminator is rather simple: it consists of just four conv-layers with the first one strided. Weights are not shared between different scale discriminators. The downsampling factor from one scale to the next is set to $\varsigma = \sqrt{2}$.

The multiscale D outputs n discrimination maps that are summed via global weighted mean pooling to yield D 's output. The weights are updated during the optimization process in a coarse-to-fine manner. Initially, the weights are such that most of the contribution to $\mathcal{L}_{\text{GAN}}$ is from the coarsest scale. As the training progresses, the contribution gradually shifts to the finer scales.

2.4. Generator Invertibility

Training G with $\mathcal{L}_{\text{GAN}}$ often leads to mode collapse where the synthesized y 's are *coherent* – the multiscale patches of y are drawn from the input image's distribution – but not *complete* – i.e. important visual information is missing from the generated y 's. To achieve better completeness, InGAN reconstructs the input image x from the output image y , ensuring no visual information was lost in y . Taking advantage of G 's automorphism allows us to re-use G to reconstruct x back from y without training an additional decoder, yielding an “Encoder-Encoder” architecture.

3. Implementation Details

We use the ADAM optimizer [19] and a linearly decaying learning rate. We train over crops, ranging from 192×192 to 256×256 , with a batch-size of 1. The default weighting of the $\mathcal{L}_{\text{reconst}}$ loss is $\lambda = 0.1$. At each iteration, parameters of a Homography transformation T are randomly sampled, resulting in different output size,

shape and aspect ratio. We employ a form of curriculum-learning so that the possible distortion allowed for T is initially very small. As the training progresses the allowed range of deformations gradually grows through the curriculum period (10k iterations) until it finally covers the entire desired range.

We employ several mechanisms for encouraging stability; spectral normalization [21] is used both in the discriminator and the generator for all layers except the last one. Batch normalization [16] is used in most conv-blocks. We also encountered a degenerate case where D was able to discriminate real patches from generated ones due to the fact that all values of the real patches were quantized to values $n/255$. To avoid this we add uniform noise in the range of $[0, 1/255]$ to the real examples before feeding them to the discriminator.

InGAN requires around 20k-75k iterations of gradient descent in order to obtain appealing results. Training takes 1.5-4 Hrs on a single V-100 GPU, regardless of the size of the input image. Once trained, InGAN can synthesize images of any desired size/shape/aspect-ratio in a single feed-forward pass. For example, InGAN can remap to VGA size (640×480) in about 40 ms (equivalent to 25 fps).

4. A Unified Framework for Multiple Tasks

InGAN provides a variety of capabilities and can be applied to multiple tasks. Moreover, it provides a unified treatment of very different data-types, ranging from pure textures to natural images, all under a single umbrella.

A single pure texture is usually captured by just a few dominant image scales. Natural images, on the other hand, tend to span a wide range of image scales, from fine-grained

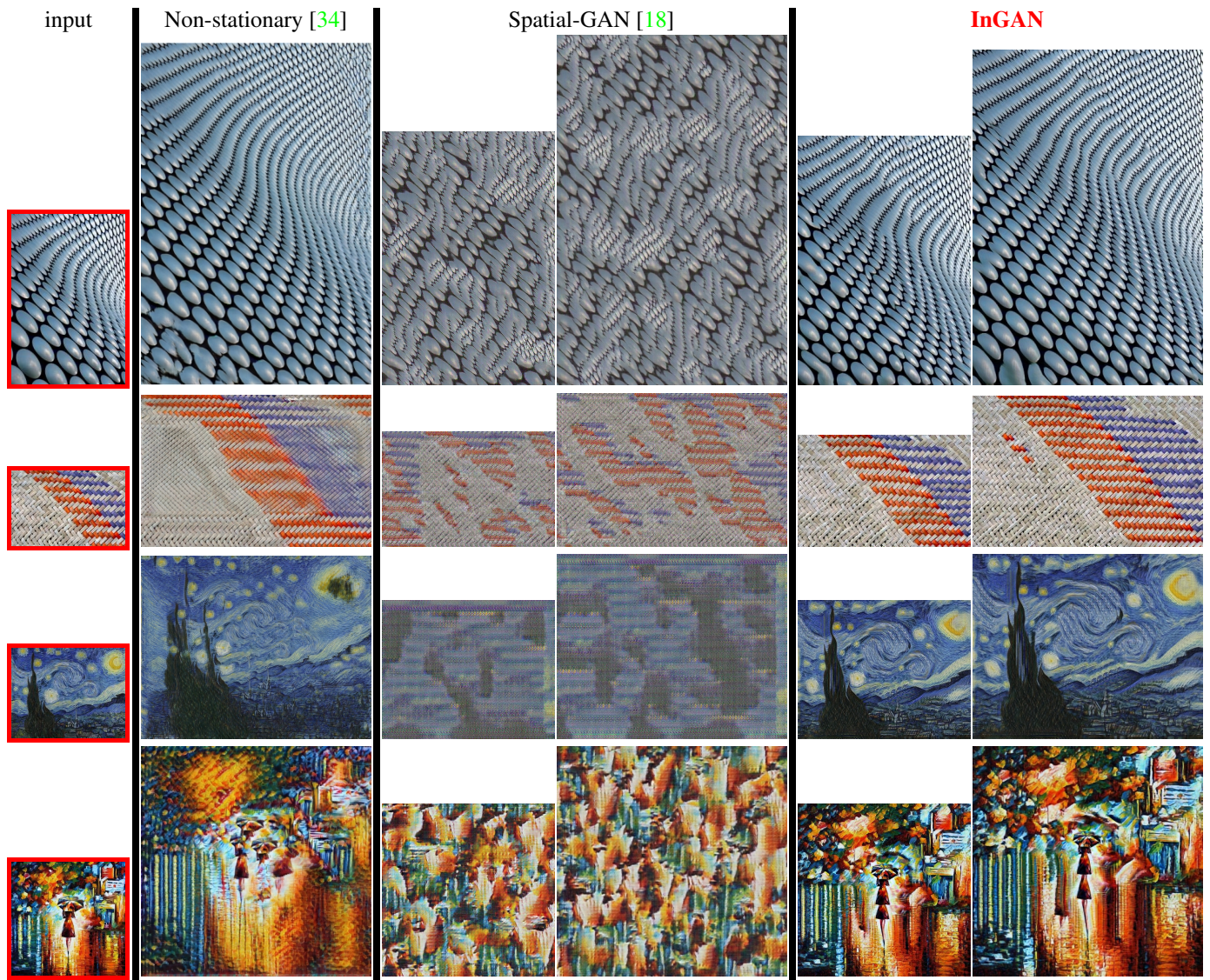


Figure 6: **Texture synthesis:** *Synthesizing textures of sizes $\times 1.5$ and $\times 2$. Note that [34] is restricted to $\times 2$. Please zoom in.*

details to coarse structures. Non-stationary textures and multi-textured images tend to reside somewhere in-between those two extremes. Each of these family of images is usually treated separately, by different specialized methods. Here we show that capturing and remapping the multiscale patch distribution of an image provides a unified treatment for all these seemingly different data-types. InGAN thus applies to a a continuum from pure textures to natural images, in a single framework.

For example, observe the corn image in Fig. 7: small image patches at fine image scales capture the tiny details of the corn seeds, while patches at coarse images scales capture the structure of an entire corn-cob. When retargeting the corn image to a wider/thinner output, entire corn-cobs are added/removed (thus matching the multiscale patch dis-

tribution). In contrast, when changing the height of the output image, small corn seeds are added/removed from each corn-cob. This multiscale patch distribution is a fundamental characteristic of both natural images and textures. Nonetheless it is important to stress that InGAN has no semantic information about “objects” or “scenes”, it only models the multiscale patch distribution of the input image.

Figs. 1,5,6,7,8 exemplify the range of capabilities and data-types handled by InGAN. Additional examples are found in the [project’s web page](#). A unique capability of InGAN is its continuous transitions between different shapes/sizes/aspect-ratios, best exemplified by the videos in the [project’s web page](#).

We next discuss a variety of tasks & capabilities provided by InGAN, all with a single network architecture. In-

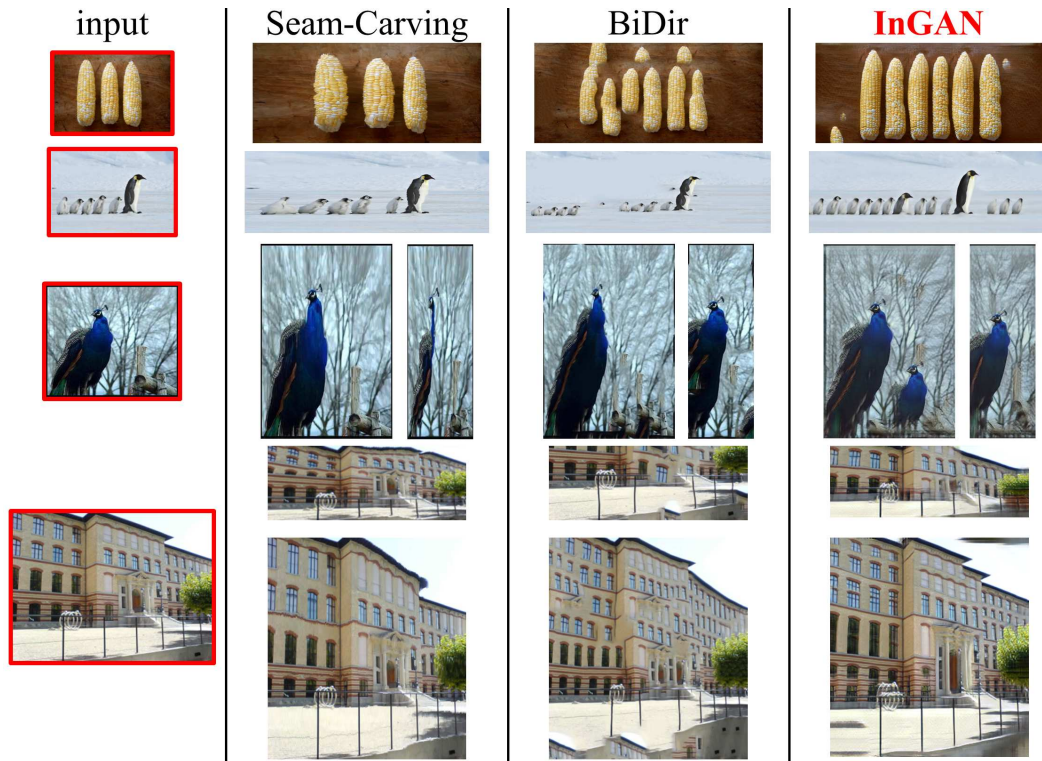


Figure 7: **Natural image retargeting:** Comparing InGAN to Bidirectional similarity [29] and Seam Carving [25]. **Please zoom in.**

GAN may not provide state-of-the-art results compared to specialized methods optimized for a specific task (often also for a specific output size). Nevertheless, InGAN compares favorably to these specialized methods, while providing a single unified framework for them all. Moreover, InGAN opens the door to new applications/capabilities.

Texture Synthesis: Texture synthesis is the task of generating novel instances of a texture that conform to some process governing the original input texture. In this work the governing process is the internal patch distribution. Gatys *et al.* [12, 11] used pretrained network features to synthesize textures. “Spatial-GAN” [18] and “Non-Stationary Texture Synthesis” [34] use a patch-based GAN in a fully convolutional manner, producing high quality textures. We refer to these kinds of textures (whether stationary or non-stationary) as **Single-texture synthesis**. Texture synthesis methods typically perform poorly on **Multi-texture synthesis** – namely, synthesizing images containing multiple textures at various scales. InGAN can handle both single- and multi-texture synthesis, thanks to its multiscale discriminator. Figs. 5 and 6 show comparisons of InGAN to specialized texture-synthesis methods, both on single- and multi-texture images ([34] is restricted to $\times 2$ outputs).

Natural Image Retargeting: Summary and Expansion

Image retargeting aims at displaying a *natural* image on a different display size, smaller or larger, often with

a different aspect ratio. Smaller representations (visual summary, thumbnail) should faithfully represent the input visual appearance as best as possible. Another goal is to generate Expanded images of the same nature (often with different aspect ratios).

There are several different notions of “image retargeting”. Some methods (e.g., [4, 33]) aim at preserving salient objects while seamlessly discarding background regions to obtain a smaller image. They mostly do smart cropping, keeping the main object centered in its original size. Some of these methods struggle when there are several dominant objects present in the image. They do not tend to perform well on images with lots of clutter and texture, nor are they catered to image expansion/synthesis. Seam-carving [1] gradually removes/adds pixel-wide “seams” that yield minimal change to image gradients. This method can handle both Summarization and Expansion.

Other methods (e.g., [29, 23]) aim at preserving local sizes/aspect-ratios of all image elements (whether salient or not) as best possible, while changing the global size/aspect-ratio of the image. They cater both Summarization and Expansion. InGAN belongs to this family of methods.

Figs. 1, 5, 7 show comparisons of InGAN to Seam-Carving and Bidirectional-Similarity, on natural images as well as non-natural ones. Since Seam-carving [1] uses local information only (it removes/adds pixel-wide “seams”), it tends to distort larger image structures under drastic

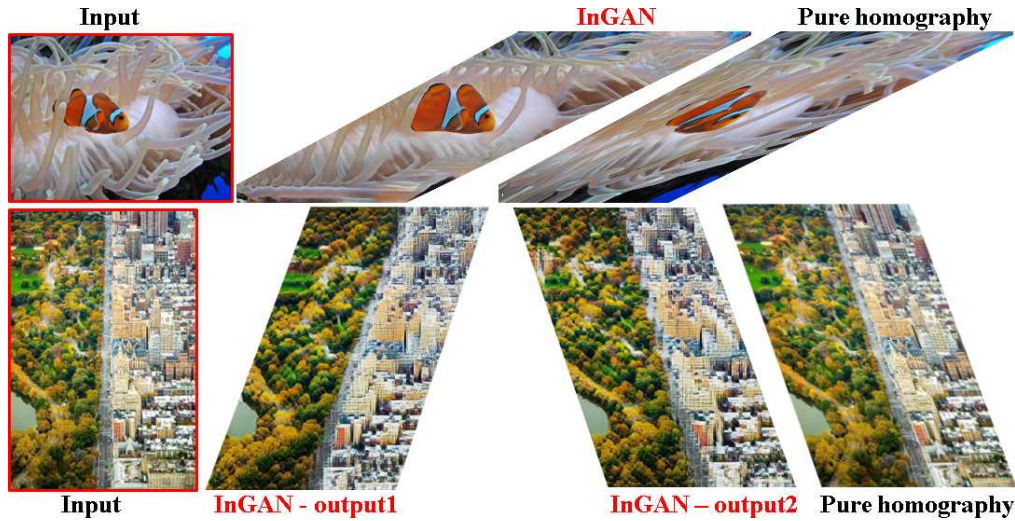


Figure 8: **Retargeting to Non-Rectangular Outputs:** *InGAN* is able to retarget to non-rectangular shapes using the geometric transformation T (e.g., homography). Note that a pure homography tilts all the elements inside the image, whereas *InGAN* preserves local shape/appearance & tilt of these elements. In particular, *InGAN* generates an illusion of retargeting to a new 3D view with correct parallax (without any 3D recovery).

changes in aspect ratio (see narrow distorted peacock in Fig. 7). Bidirectional-Similarity [29] handles this by using image patches at various scales, but requires solving a new optimization problem for each output size/aspect-ratio. In contrast, *InGAN* synthesizes a plethora of new target images of different shapes with a single trained network.

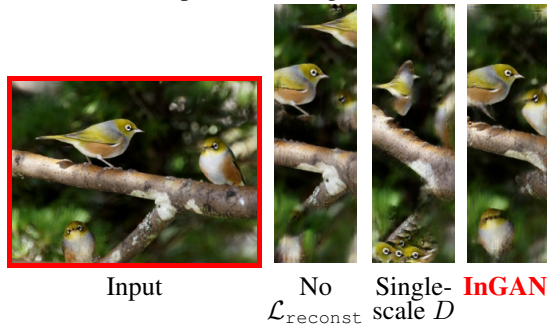


Figure 9: **Ablation study:** Omitting $\mathcal{L}_{reconst}$ or using a single-scale D , degrades the results compared to full *InGAN* architecture.

Image Retargeting to Non-rectangular Shapes: Unlike previous methods, *InGAN* retargets images into non-rectangular outputs. This is made possible by introducing random invertible geometric transformations in *InGAN*’s generator. Our current implementations uses 2D projective transformations, but the framework permits any invertible transformation. Figs. 1 and 8 display a few such examples. Note that a pure homography tilts all the elements inside the image. In contrast, *InGAN* preserves local shape & tilt of these elements despite the severe change in global shape of the image. In particular, while the synthesized visual quality is not very high under extreme shape distortions, *InGAN* generates an interesting illusion of retargeting into a new 3D view with correct parallax (but without any 3D estimation).



Figure 10: **Failure example:** *Input* in red. *InGAN* has no semantic understanding of “objects” or “scenes”, it only models the multiscale patch distribution of the input image, hence cannot distinguish between object-parts and entire objects.

5. Ablation Study and Limitations

We conducted an ablation study to verify the importance of: (i) the “encoder-encoder” architecture with its ℓ_1 reconstruction loss, and (ii) the importance of multiple scales in the discriminator D . Fig. 9 shows one such example: Training *InGAN* without $\mathcal{L}_{reconst}$ (left-most result) shows unstructured output: two birds are completely missing and the dominant bird is split into two. Using a single scale D (middle result) makes G generate a result that is locally coherent, but lacks large scale structures. The birds were completely destroyed. In contrast, the full *InGAN* (right-most) with $\mathcal{L}_{reconst}$ and multiscale D maintains both fine details and coarse structures. In particular, all 3 birds are in the output.

Limitations: *InGAN* is unsupervised – it has no additional information other than the input image. In particular, *InGAN* has no semantic understanding, no notion of “objects” or “scenes”. Its sole objective is capturing and remapping the multiscale patch distribution of the input image. Thus, *InGAN* sometimes produces funny/unnatural results. Fig 10 shows such an example: *InGAN* produces an output that is both *coherent* and *complete* (all local elements are preserved), yet is incorrect in its semantic meaning.

References

- [1] Shai Avidan and Ariel Shamir. Seam carving for content-aware image resizing. *ACM Trans. Graph.*, 26(3), July 2007. 1, 7
- [2] Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan B. Goldman. Patchmatch: A randomized correspondence algorithm for structural image editing. In *SIGGRAPH*, 2009. 1
- [3] Antoni Buades, Bartomeu Coll, and J-M. Morel. A non-local algorithm for image denoising. In *CVPR*, volume 2, pages 60–65, 2005. 1
- [4] Donghyeon Cho, Jinsun Park, Tae-Hyun Oh, Yu-Wing Tai, and In So Kweon. Weakly and self-supervised learning for content-aware deep image retargeting. In *The IEEE International Conference on Computer Vision (ICCV)*, 2017. 1, 7
- [5] Taeg Sang Cho, Shai Avidan, and William T. Freeman. The patch transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010. 1
- [6] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007. 1
- [7] Vincent Dumoulin and Francesco Visin. A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285*, 2016. 4
- [8] Alexei Efros and Thomas Leung. Texture synthesis by non-parametric sampling. In *ICCV*, volume 2, pages 1033–1038, 1999. 1
- [9] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, 2006. 1
- [10] Tomer Galanti, Lior Wolf, and Sagie Benaim. The role of minimal complexity functions in unsupervised learning of semantic mappings. *arXiv preprint arXiv:1709.00074*, 2017. 4
- [11] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. A neural algorithm of artistic style, 2015. 7
- [12] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. Texture synthesis using convolutional neural networks, 2015. 7
- [13] Daniel Glasner, Shai Bagon, and Michal Irani. Super-resolution from a single image. In *International Conference on Computer Vision (ICCV)*, 2009. 1
- [14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014. 3
- [15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015. 4
- [16] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37, ICML’15*, pages 448–456. JMLR.org, 2015. 5
- [17] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 4
- [18] Nikolay Jetchev, Urs Bergmann, and Roland Vollgraf. Texture synthesis with spatial generative adversarial networks, 2016. 3, 5, 6, 7
- [19] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. 5
- [20] Xudong Mao, Qing Li, Haoran Xie, Raymond Y. K. Lau, and Zhen Wang. Least squares generative adversarial networks. In *Computer Vision (ICCV), IEEE International Conference on*, 2017. 4
- [21] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations*, 2018. 5
- [22] Augustus Odena, Vincent Dumoulin, and Chris Olah. Deconvolution and checkerboard artifacts. *Distill*, 2016. 4
- [23] Yael Pritch, Eitam Kav-Venaki, and Shmuel Peleg. Shift-map image editing. In *2009 IEEE 12th International Conference on Computer Vision*, pages 151–158. IEEE, 2009. 1, 7
- [24] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention MICCAI 2015*, page 234241, 2015. 4
- [25] Michael Rubinstein. MATLAB reimplementation of seam carving. http://people.csail.mit.edu/mrub/code/seam_carving-1.0.zip, 2009. 5, 7
- [26] Tamar Rott Shaham, Tali Dekel, and Tomer Michaeli. SinGAN: Learning generative model from a single natural image. In *ICCV*, 2019. 3
- [27] Assaf Shocher, Shai Bagon, Phillip Isola, and Michal Irani. Internal distribution matching for natural image retargeting. In *arxiv 1812.00231v1*, 2018. 3
- [28] Assaf Shocher, Nadav Cohen, and Michal Irani. Zero-shot super-resolution using deep internal learning. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 3
- [29] Denis Simakov, Yaron Caspi, Eli Shechtman, and Michal Irani. Summarizing visual data using bidirectional similarity. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008. 1, 4, 5, 7, 8
- [30] Dmitry Ulyanov, Vadim Lebedev, Andrea Vedaldi, and Victor S Lempitsky. Texture networks: Feed-forward synthesis of textures and stylized images. In *ICML*, volume 1, page 4, 2016. 3
- [31] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9446–9454, 2018. 3

- [32] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 4
- [33] Lior Wolf, Moshe Guttman, and Daniel Cohen-Or. Non-homogeneous content-driven video-retargeting. In *IEEE 11th International Conference on Computer Vision*, 2007. 1, 7
- [34] Yang Zhou, Zhen Zhu, Xiang Bai, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. Non-stationary texture synthesis by adversarial expansion. *ACM Trans. Graph.*, 37(4):49:1–49:13, July 2018. 3, 5, 6, 7
- [35] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networkss. In *Computer Vision (ICCV), IEEE International Conference on*, 2017. 3
- [36] Maria Zontak and Michal Irani. Internal statistics of a single natural image. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011. 1