



This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail.

Heljakka, Ari; Solin, Arno; Kannala, Juho

Recursive Chaining of Reversible Image-to-Image Translators for Face Aging

Published in: Advanced Concepts for Intelligent Vision Systems - 19th International Conference, ACIVS 2018, Proceedings

DOI: 10.1007/978-3-030-01449-0_26

Published: 01/01/2018

Document Version Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:

Heljakka, A., Solin, A., & Kannala, J. (2018). Recursive Chaining of Reversible Image-to-Image Translators for Face Aging. In *Advanced Concepts for Intelligent Vision Systems - 19th International Conference, ACIVS 2018, Proceedings* (pp. 309-320). (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 11182 LNCS). Springer. https://doi.org/10.1007/978-3-030-01449-0_26

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Recursive Chaining of Reversible Image-to-Image Translators for Face Aging

Ari Heljakka^{1,2}, Arno Solin², and Juho Kannala²

¹ GenMind Ltd, Finland

² Department of Computer Science, Aalto University, Espoo, Finland {ari.heljakka,arno.solin,juho.kannala}aalto.fi

Abstract. This paper addresses the modeling and simulation of progressive changes over time, such as human face aging. By treating the age phases as a sequence of image domains, we construct a chain of transformers that map images from one age domain to the next. Leveraging recent adversarial image translation methods, our approach requires no training samples of the same individual at different ages. Here, the model must be flexible enough to translate a child face to a young adult, and all the way through the adulthood to old age. We find that some transformers in the chain can be recursively applied on their own output to cover multiple phases, compressing the chain. The structure of the chain also unearths information about the underlying physical process. We demonstrate the performance of our method with precise and intuitive metrics, and visually match with the face aging state-of-the-art.

Keywords: Deep Learning \cdot Transfer Learning \cdot GAN \cdot Face Synthesis \cdot Face Aging

1 Introduction

Generative Adversarial Network (GAN) [8] variants have been successful for various image generation and transformation tasks. For image-to-image translation (such as mapping sketches to photographs), they have achieved state-of-the-art results, with paired training data [9] and without it [13,28,27,17].

This paper generalizes image-to-image mapping to a sequential setting. Previous works have not focused on recursive application of the models on their own outputs, and there have been no extensions to apply the method for a sequence of domains, even though, *e.g.*, [5] allow applying several *different kinds* of domain transformations to the same image. We propose a recursive adversarial domain adaptation approach that is capable of producing step-wise transformations for human aging, as visualized in the examples in Fig. 1. We use the *reversible* image translation approach of [13,28,27].

Previously, Antipov *et al.* [3] applied a conditional GAN to simulating human face aging with good initial results. Their model distinguishes itself by explicitly enforcing the preservation of identity in face transformations. However, in reversible transformers such as ours, the preservation of identity requires



Fig. 1: Examples of face aging transformations. Row 1: Non-recursive transformation from 15-year-old original to 65-year-old synthetic. Row 2: Partially recursive transformation by re-using the transformer of $25 \rightarrow 35$ for also $35 \rightarrow 45$. Row 3: Non-recursive transform of an approximately 55-year-old original to older (65-year-old) and younger (towards 15-year-old). Row 4: Partially recursive transformation by re-using the transformer of $35 \rightarrow 25$ for also $45 \rightarrow 35$.

no explicit measures. By learning a reversible mapping between two faces, the transformer is inclined to preserve the identity information. Face aging has been tackled before in approaches such as [25,12,7,16]. One can also find adjustable parameters to modulate the prevalence of general attributes in an image [15]. While learning adjustable knobs allows for fine-grained control, our approach is suitable for robust learning of pronounced sequential changes.

The rationale of this paper follows from modeling progressive changes. Consider the development of an entity that can be approximated as a closed system (e.g., an aging human face, a deteriorating surface, a growing tree, or a changing city outline). Visual representation of such development can be approximated as a succession of snapshots taken over time, so that one snapshot is translated to the next one, then to the next one,*etc.*Such a translation requires Markovian development, so that each image carries sufficient information to enable translating it to the next stage. This allows for*e.g.*face aging, but not for modeling the day–night cycle.

Utilizing unpaired image-to-image translation methods, we can learn these translation steps. Suppose we have training samples from an entity class P such

that we can at least roughly assign each one to a bin P_i that corresponds to its developmental phase *i*. Knowing the ordering of the phases, the model can learn to transform an image from stage P_i to stage P_{i+1} . The proposed method can simulate one such possible line of development, with the hope that the real underlying process has sufficient regularity to it. In this paper, the application of interest is human face aging, which is both a challenging modeling problem and the results are easy to validate by human readers (as humans are highly specialized in face perception).

The contributions of this paper are as follows.

- We show how a chain framework of unsupervised reversible GAN transformers can be constructed to convert a human face into a desired age, along with the necessary image pre-processing steps.
- We show that using an auxiliary scoring method (*e.g.* a face age estimator) to rank the transformers between training stages, we can leverage transfer learning with a simple meta-training scheme that re-uses the best transformers and compresses the chain with no performance cost. The auxiliary score provides no learning gradient for an individual transformer. However, we show that the score does consistently reflect the training progress.
- Consequently, our ability to compress the transformer chain can be used to separate the high-level linear (*e.g.* increasing formation of wrinkles) and nonlinear stages of development, producing a compact human-interpretable development descriptor.

2 Proposed Method

2.1 Sequential transformations

In reversible GAN transformers, such as [28], there are two generator networks and two discriminator networks. One generator transforms images from domain A to domain B and the other from B to A. One discriminator separates between whether an image originates from the training set of B or from the first generator. The other discriminator does the same for A and the second generator. The resulting generators will have learnt to map image features between the domains.

We can treat this kind of a network as a single building block of a chain of domain transformers. For paired domains such as summer-winter or horseszebras, the notion of such chaining does not apply. However, it does apply for classes of closed systems that evolve over time along a path of certain regularity. If we can disentangle the relevant patterns of development, a transition between two stages may occur by repeated or one-time application of those patterns.

A sequential transformation can be described as a composition of operators, with fundamental relations such as reflexivity and transitivity. Given a mapping F from domain A to domain B, what is the intended meaning of F(F(a)), $a \in A$? In the context of regular non-sequential transformations such as [28], a reasonable option would be to require either reflexivity as $F(F(a)) \in A$ or antireflexivity as $F(F(a)) \notin A$ or $F(F(a)) \in B$. These constraints could be added to existing image-to-image translators. The more relevant case here is transitivity. For domains A_t , $t = \{1, 2, 3\}$, and $a_i \in A_i$, we would like to have $F(a_1) \in A_2$, and also $F(F(a_1)) \in A_3$. This allows us to enforce the transformer's ability to perform multiple transformations, turning it into an extra loss term. One may also vary the 'domain thickness' with respect to F, *i.e.* the measure of how many times F should be applied in order to move from domain *i* to domain i + 1 (*e.g.* 2, 1, 1/2, *etc.*). With enough domains, thickness of 1 suffices.

2.2 Meta-training of transformer models

We train a chain of transformers to cover each developmental stage, as given in Alg. 1. The algorithm requires a way to measure model performance at metalevel, i.e. between the actual training runs. For GANs, traditional likelihood measures are problematic [24,26] and in some cases inversely correlated with image quality [6]. GANs are often prone to collapse into a specific mode of the distribution, missing much of the diversity of the training data. Trainable auxiliary evaluators like [6,18] do help, but for many application domains, we already have automated deterministic scoring tools for images. For human face aging, such auxiliary evaluators exist. We chose [2].

The algorithm starts by training a separate transformer network Φ_1 between the first two consequtive datasets, *e.g.* 15 and 25-year-old face images, for Nsteps (N determined empirically). Transformers trained in this way are called *baseline*. We use CycleGAN [28], but the algorithm is applicable for any imageto-image transformer.

Then, we train another model Φ_2 in the exact same way for the next stage pair, e.g. mapping from 25 to 35-year-olds. But now, we also re-train a copy of the model Φ_1 recycled from the previous stage with additional N/2 steps on its earlier data and N/2 on the next stage data. For example, Φ_1 , originally trained for 15 \rightarrow 25 conversion, is now re-trained also on 25 \rightarrow 35 data. In order to benchmark Φ_1 and Φ_2 , we measure the perceived age of the faces in the images they produce, with the estimator [2]. The transformations should age by 10 years, so we can score each transformer simply by the error in the mean age of its output distribution, normalized by its standard deviation (alg. lines 8–9).

If the re-trained transformer is not significantly worse on this new transformation than the baseline transformer, we discard the baseline transformer and replace it with the re-trained transformer. Otherwise we discard the retrained one. The remaining transformer will then be tried out on the next stage, and continuing until the end of the transformer chain. By virtue of reversible transformers, we could also run the same algorithm backwards, finding the best models that make an old face young again (rows 2 and 4 in Fig. 1).

The auxiliary scoring method is only used for a binary choice between two competing transformers during stage transitions. The score, therefore, provides no gradient for the training. Furthermore, our scoring model has been trained on an altogether different dataset and model architecture. Retrospectively, we calculate what the scores would have been during training (Fig. 2b).

Alg. 1 Greedy forward-mode recursive transformer chain with two-step backward-compatibility. In the experiments, we used $\mu_i^{\text{target}} = [15, 25, \dots, 65],$ $N_D = 6, S = 600,000, \varepsilon = 0.1.$

1: Require: Number of stages N_D , data sets D_i and target mean age μ_i^{target} with $i \in [1, N_D]$, trainable models $\Phi_j : \mathbb{R}^{256 \times 256} \to \mathbb{R}^{256 \times 256}$ with $j \in [1, N_D - 1]$, number of steps S, auxiliary age estimator $\Gamma : \mathbb{R}^{256 \times 256} \to \mathbb{R}$ 2: Initialize: $a \leftarrow 1$ \triangleright Denote the index of the Φ model we try to re-use $\Phi_1 \leftarrow \operatorname{train}(\Phi_1, [(D_1, D_2)], S)$ 3: 4: for $i = 2, \ldots, (N_D - 1)$ do $\Phi_i \leftarrow \operatorname{train}(\Phi_i, [(D_i, D_{i+1})], S)$ 5: $\Phi'_a \leftarrow \operatorname{copy}(\Phi_a)$ 6: $\Phi'_a \leftarrow \operatorname{train}(\Phi'_a, [(D_{i-1}, D_i), (D_i, D_{i+1})], S)$ 7:
$$\begin{split} & E' \leftarrow \left| \mathbb{E}_{d \sim D_i} [\| \Gamma(\Phi'_a(d)) - \mu^{target}_{i+1} \|_1] / \sigma(\Gamma(\Phi'_a(d))) \right| \\ & E \leftarrow \left| \mathbb{E}_{d \sim D_i} [\| \Gamma(\Phi_i(d)) - \mu^{target}_{i+1} \|_1] / \sigma(\Gamma(\Phi_i(d))) \right| \\ & E \leftarrow \left| \mathbb{E}_{d \sim D_i} [\| \Gamma(\Phi_i(d)) - \mu^{target}_{i+1} \|_1] / \sigma(\Gamma(\Phi_i(d))) \right| \end{split}$$
8: 9: 10: if $|E - E'| < \varepsilon$ then \triangleright Recycled model wins 11: $release(\Phi_i)$ 12: $release(\Phi_a)$ $\Phi_a \leftarrow \Phi_i \leftarrow \Phi'_a$ 13: \triangleright Upgrade Φ_a and try re-using again 14: else 15: $release(\Phi'_a)$ 16: $a \leftarrow i$ \triangleright Next, try re-using the most recent base model

The algorithm uses a simple greedy search. By enforcing backward compatibility only to the previous transformation stage, we will only prevent forgetting in short sequences. For the six domains, we found the incurring performance penalty to be minor (Fig. 2b). To scale up, we could simply cap the maximum number of re-uses of a model or the allowed error from forgetting, or add equal parts of even earlier domains to the training set of the re-used model.

3 Experiments

3.1 Dataset and auxiliary age estimator

For training, we use the Cross-Age Celebrity Dataset (CACD, [4]), with a large number of age-annotated images. Our chosen auxiliary age estimator [2], pretrained on the IMDB-Wiki dataset [21], was used as-is (the estimator has not seen any of our training data). We found it necessary to improve the CACD data in two ways. First, we used the off-the-shelf face alignment utility of [22] to crop and align the faces based on landmarks. Second, despite the existing age annotations of CACD, they are not accurate enough for age-estimation [1]. We ran the auxiliary age estimator on the data and found major discrepancy between the annotations and the results of the estimator. We also confirmed visually that the estimator was closer to the ground truth. Using the estimator as ground-truth, we re-annotated the whole CACD data accordingly. Our model, however, has access neither to the age information nor the estimator. It only knows that the pictures come from different domains, at 10-year accuracy. For validation, we use a small subset of the IMDB-Wiki dataset.

We divided the data into slots 2–18, 19–29, 30–39, 40–49, 50–59, and 60–78 (six domains, with five direct transformation paths). This enables direct comparison with [3]. We removed enough samples so that the mean age in the sets is 10 years apart—15, 25, 35, 45, 55, and 65, respectively. Since the numbers of samples may differ between domains, we measure training progress by the number of steps, not epochs, so as to maintain commensurability between stages.

3.2 Architecture and training

The architecture of each transformer module follows [28] (which re-uses structures from [10] and [9]), with two generators and two discriminators per module. The generators utilize convolution layers for encoding, 9 ResNet blocks for transformation, and deconvolution layers for decoding (original source code adopted from [19]). Our image preprocessing reduces the real face resolution to 132×132 , which is then upscaled to 256×256 for computational efficiency using off-the-shelf Mitchell–Netravali filter [20]. The full chain is composed of five independently trained successive modules, so that one module feeds its output to the next. The results of each transformation stage can be evaluated independently, and external input can be fed in at any stage of the chain. For recursion, one can co-train with 3 stages (*e.g.*15 \rightarrow 25 and 25 \rightarrow 35), or apply the model twice (15 \rightarrow 35).

We trained with ADAM ([14], learning rate 0.0002, $\beta_1 = 0.5$, $\beta_2 = 0.999$) with a batch size one. As in [28,23], we update the discriminators using a buffer of 50 recently generated images rather than only the most recently produced ones. The training time on an NVIDIA P100 workstation was 540 h (300 h for the baseline, 60 h for each re-training session).

3.3 Human face age progression

Our solution is based on a pipeline of successive transformations. In order to train a single transformation, say $25\rightarrow 35$, we train a single transformer network. By the reversible architecture, we simultaneously train the network to carry out the full inverse transformation, $35\rightarrow 25$.

We trained and evaluated each transformer network according to Alg. 1. For human face aging, our hypothesis was that the network trained for transforming faces from $15\rightarrow 25$ would not generalize to the other stages, whereas the network that transforms from $25\rightarrow 35$ would generalize to multiple later stages as well. We confirm this both during training-time (Fig. 2a) and validation (Fig. 2b). The algorithm drops the baseline transformer $35\rightarrow 45$ off the chain ($\epsilon > 0.35$ would drop the next one, too). From this result, we can read off the development descriptor of the form $F(\text{age} = 65) = F_{55\rightarrow 65}(F_{45\rightarrow 55}(F_{25\rightarrow 35}^2(F_{15\rightarrow 25}(\text{age} = 15))))$. We expected that if we can re-use some transformers, this provides us with highlevel information about the underflying development pattern. While our result alone only gives tentative support to this idea, we expect this approach to be also applicable with more fine-grained precision.



Fig. 2: (a) Performance of modules as a function of training steps (with stagespecific transformation target), greedily selecting the best-performing models for re-use. The error bars indicate the standard deviation of the ages of generated test samples. Some of the variance is by design, as we intentionally challenge the model with a wide *range* of ages for robustness and generalization in training data (for clarity of visualization, the ground truth variance not shown). (b) Resulting age after transformation in validation set, as a function of the input age. During training, the model learned that $25 \rightarrow 35$ transformer can be recycled for $35 \rightarrow 45$ to transform better than the freshly trained $35 \rightarrow 45$ transformer. The $35 \rightarrow 45$ transformer was thus discarded (see Alg. 1). $15 \rightarrow 25$ and $45 \rightarrow 55$ transformers succeeded on their baseline but were not re-usable.

3.4 Comparison

In Fig. 1, we show sample transformation paths, with a 15-year-old successively transformed to 65, and a 55-year-old transformed back to 15 and forward to 65. The results are visually at least at the level of [3]. Additional results are shown in Figures 3 and 4.

As most methods are concerned with singular transformations, our most relevant points of comparison are [3] and [15]. However, both show results where several successive stages of development show effectively no changes at all, and no age measurements. Our results show greater variation between each stage, with not too many artifacts even at relatively low resolution. The improvements are presumably due to both the differences between reversible GANs and conditional GANs, and to the use of the transformer chain. The chain, of course, comes with additional computational cost.

In average, our compressed chain produced the target age in the validation set with an error less than 4.5% (maximum of 3.2 years) and standard deviations between 2.4–5.6 in different age groups. The performance drop from the baseline due to compression was negligible. [3] reports an estimation accuracy 17% lower for synthetic images than natural ones. This may reflect the improvements in visual accuracy and variation in our model.

4 Discussion and Conclusion

In this paper, we showed that a transformer chain composed of reversible GAN image transformer modules can learn a complex multi-stage face transformation task. The domain-agnostic base algorithm is expected to generalize to other kinds of temporal progression problems. Notably, we found that a single transformer can carry out the face transformations $25\rightarrow35$ and $35\rightarrow45$ (and, with minor loss of accuracy, also $45\rightarrow55$). One might suspect this to be because not many visible changes happen during this time. However, the auxiliary age estimator can still discern the ages 25, 35, and 45 easily. As the estimator and transformer chains are independent, this indicates the presence of real changes that the networks capture in a systematic way, even when they are hard for humans to discern.

On some problem domains, the full chain may be relatively long and slow to train. Also, there is currently little control over the transformation paths, and the chain may not yield a range of varied outcomes in contexts where the developmental paths could diverge in different directions. Recent methods for improving the resolution of GAN-generated images [11] could be combined with our method. Follow-up work should evaluate the extent to which the layers of separate transformer modules can be shared, so as to reduce the total training time. More comprehensive evaluation of the models in both directions would likely find more compact chains. In semi-supervised setting, using a small number of paired examples would likely improve our results.

The code for replicating the results is available online: <code>https://github.com/AaltoVision/img-transformer-chain.</code>



Fig. 3: Examples of transforming an approximately 15-year-old to 65-year-old. For each identity, row 1 shows the non-recursive transformation (applying the baseline transformer on each stage). Row 2 shows the partially-recursive transformation, with the double-trained $25 \rightarrow 35$ transformer applied also to $35 \rightarrow 45$ (the best chain, according to Alg. 1). Row 3 shows the transformation with the most recursive steps, with the triple-trained $25 \rightarrow 35$ transformer applied also to both $35 \rightarrow 45$ (the most compact chain, picked manually).



Fig. 4: Examples of transforming an approximately 55-year-old to older (65-year-old) and younger (towards 15-year-old). For each identity, row 1 shows the non-recursive transformation (applying the baseline transformer on each stage). Row 2 shows the partially-recursive transformation, with the double-trained $35\rightarrow 25$ transformer applied also to $45\rightarrow 35$ (the best chain, according to Alg. 1). Row 3 shows the transformation with the most recursive steps, with the triple-trained $35\rightarrow 25$ transformer applied also to both $45\rightarrow 35$ and $55\rightarrow 45$ (the most compact chain, picked manually).

Acknowledgments

This research was supported by GenMind Ltd and the Academy of Finland grants 308640, 277685, and 295081. We acknowledge the computational resources provided by the Aalto Science-IT project.

References

- 1. Cross-age reference coding for age-invariant face recognition and retrieval. http://bcsiriuschen.github.io/CARC/, accessed: 2018-05-18
- Antipov, G., Baccouche, M., Berrani, S., Dugelay, J.: Apparent age estimation from face images combining general and children-specialized deep learning models. In: CVPR Workshops (2016)
- Antipov, G., Baccouche, M., Dugelay, J.L.: Face aging with conditional generative adversarial networks. In: ICIP (2017)
- 4. Chen, B.C., Chen, C.S., Hsu, W.H.: Cross-age reference coding for age-invariant face recognition and retrieval. In: ECCV. pp. 768–783 (2014)
- Choi, Y., Choi, M., Kim, M., Ha, J.W., Kim, S., Choo, J.: StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. arXiv preprint arXiv:1711.09020 (2017)
- Danihelka, I., Lakshminarayanan, B., Uria, B., Wierstra, D., Dayan, P.: Comparison of maximum likelihood and GAN-based training of real NVPs. arXiv preprint arXiv:1705.05263 (2017), https://arxiv.org/abs/1705.05263
- Duong, C., Quach, K., Luu, K., Savvides, M.: Temporal non-volume preserving approach to facial age-progression and age-invariant face recognition. In: ICCV (2017)
- Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. In: NIPS (2014)
- Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: CVPR (2017)
- 10. Johnson, J., Alahi, A., Li, F.: Perceptual losses for real-time style transfer and super-resolution. In: ECCV (2016)
- Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of GANs for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196 (2017)
- Kemelmacher-Shlizerman, I., Suwajanakorn, S., Seitz, S.: Illumination-aware age progression. In: CVPR (2014)
- Kim, T., Cha, M., Kim, H., Lee, J.K., Kim, J.: Learning to discover cross-domain relations with generative adversarial networks. In: ICML. PMLR, vol. 70, pp. 1857– 1865 (2017)
- Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: ICLR (2015), https://arxiv.org/abs/1412.6980
- 15. Lample, G., Zeghidour, N., Usunier, N., Bordes, A., Denoyer, L., Ranzato, M.: Fader networks: Manipulating images by sliding attributes. In: NIPS (2017)
- Lee, M., Seok, J.: Controllable generative adversarial network. arXiv preprint arXiv:1708.00598 (2017)
- Liu, M.Y., Breuel, T., Kautz, J.: Unsupervised image-to-image translation networks. In: NIPS (2017)
- Lopez-Paz, D., Oquab, M.: Revisiting classifier two-sample tests. In: ICLR (2017), https://arxiv.org/abs/1610.06545

- 19. LynnHo (GitHub user): CycleGAN Tensorflow PyTorch. https://github.com/ LynnHo/CycleGAN-Tensorflow-PyTorch-Simple, gitHub repository
- Mitchell, D.P., Netravali, A.N.: Reconstruction filters in computer-graphics. ACM Siggraph Computer Graphics 22(4), 221–228 (1988)
- 21. Rothe, R., Timofte, R., Gool, L.V.: DEX: Deep EXpectation of apparent age from a single image. In: ICCV, Looking at People Workshop (2015)
- 22. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: CVPR (2015)
- Shrivastava, A., Pfister, T., Tuzel, O., Susskind, J., Wang, W., Webb, R.: Learning from simulated and unsupervised images through adversarial training. In: CVPR (2017), https://arxiv.org/abs/1612.07828
- Theis, L., van den Oord, A., Bethge, M.: A note on the evaluation of generative models. In: ICLR (2016), https://arxiv.org/abs/1511.01844
- Tiddeman, B., Burt, M., Perrett, D.: Prototyping and transforming facial textures for perception research. IEEE Computer Graphics and Applications 21(5), 42—-50 (2001)
- Wu, Y., Burda, Y., Salakhutdinov, R., Grosse, R.: On the quantitative analysis of decoder-based generative models. In: ICLR (2017), https://arxiv.org/abs/1611. 04273
- 27. Yi, Z., Zhang, H., Tan, P., Gong, M.: DualGAN: Unsupervised dual learning for image-to-image translation. In: ICCV (2017)
- Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: ICCV (2017)