Analogical Image Translation for Fog Generation

Rui Gong, [1] **Dengxin Dai,** [1] **Yuhua Chen,** [1] **Wen Li,** [3] **Danda Pani Paudel,** [1] **Luc Van Gool** [1,2]
[1]Computer Vision Lab, ETH Zurich
[2]VISICS, KU Leuven
[3]University of Electronic Science and Technology of China
{gongr, dai, yuhua.chen, paudel, vangool}@vision.ee.ethz.ch, liwenbnu@gmail.com

Abstract

Image-to-image translation is to map images from a given *style* to another given *style*. While exceptionally successful, current methods assume the availability of training images in both source and target domains, which does not always hold in practice. Inspired by humans' reasoning capability of analogy, we propose analogical image translation (AIT) that exploit the concept of *gist*, for the first time. Given images of two styles in the source domain: $\mathcal{A}$ and $\mathcal{A}'$, along with images $\mathcal{B}$ of the first style in the target domain, learn a model to translate $\mathcal{B}$ to $\mathcal{B}'$ in the target domain, such that $\mathcal{A} : \mathcal{A}' :: \mathcal{B} : \mathcal{B}'$. AIT is especially useful for translation scenarios in which training data of one style is hard to obtain but training data of the same two styles in another domain is available. For instance, in the case from normal conditions to extreme, rare conditions, obtaining real training images for the latter case is challenging. However, obtaining synthetic data for both cases is relatively easy. In this work, we aim at adding adverse weather effects, more specifically fog, to images taken in clear weather. To circumvent the challenge of collecting real foggy images, AIT learns the *gist* of translating synthetic clear-weather to foggy images, followed by adding fog effects onto real clear-weather images, without ever seeing any real foggy image. AIT achieves zero-shot image translation capability, whose effectiveness and benefit are demonstrated by the downstream task of semantic foggy scene understanding.

Introduction

Image-to-image translation has enjoyed tremendous progress in the last years. Excellent methods have been developed for a diverse set of learning paradigms such as supervised (Isola et al. 2017), unsupervised (Zhu et al. 2017; Huang et al. 2018) and few-shot (Liu et al. 2019). While exceptionally successful, current methods have a shared assumption that training data, be it paired or unpaired, is available for both *styles* [1]. This may limit the use of image translation when data in one of the two *styles* is hard to obtain, e.g. translation from a normal condition to an extreme, corner-case condition. To address this, we take a new route and propose analogical image translation (AIT) which learns image translation via analogy.

[1]We reserve 'domains' for analogy since the sought analogy exists between domains, and use 'styles' for image translation.

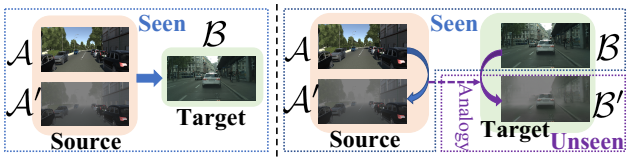
Analogy is a basic reasoning process to transfer information or meaning from the source to the target. Humans use it commonly to solve problems, provide explanations and make predictions (Hertzmann et al. 2001; Schunn and Dunbar 1996). In this paper, we explore the use of analogy as a means for extracting the *gist* of image translation in the source domain and apply it the target domain. Particularly, we aim to solve the following problem:

Problem ("Analogical Image Translation"): Given images of two styles in the source domain: $\mathcal{A}$ and $\mathcal{A}'$, with images $\mathcal{B}$ of the first style in the target domain, learn the *translation gist* and apply it to $\mathcal{B}$ to obtain $\mathcal{B}'$, such that $\mathcal{A} : \mathcal{A}' :: \mathcal{B} : \mathcal{B}'$.

The above problem cannot be addressed by the traditional image translation methods, due to the absence of $\mathcal{B}'$. On the other hand, there exist only one work, up to our knowledge, that exploits the concept of analogy for deep image translation, namely (Chen, Xu, and Jia 2020). However, (Chen, Xu, and Jia 2020) does not use the concept of *gist*. In this work, we demonstrate that the task of AIT can greatly benefit from modeling the concept of *gist*. In fact, our work also introduces the formal concept of *gist* for the task at hand.

A schematic comparison of AIT to the standard image translation is shown in Fig. 1. Our work is partially motivated by the difficulty in obtaining real training images for semantic understanding tasks of autonomous driving in adverse conditions, *e.g.*, the foggy weather. Despite tremendous progress, prior works in semantic scene understanding (Ronneberger, Fischer, and Brox 2015; Chen et al. 2017; Yu and Koltun 2016; Zhao et al. 2017; Lin et al. 2017) have mostly focused on the clear-weather, leading to unsatisfactory performance for adverse conditions (Halder, Lalonde, and Charette 2019; Sakaridis, Dai, and Van Gool 2018; Blum et al. 2019; Li et al. 2017). Collecting large-scale training datasets for such adverse conditions, and other corner cases, may resolve the issue. Unfortunately, such solutions are neither scalable and affordable, nor very practical.

To address the issues of scarce data, recent works focus on synthesizing fog effects onto clear-weather images by using a physical optical model (Sakaridis, Dai, and Van Gool 2018; Hahner et al. 2019; Ren et al. 2016). The success of these methods hinges on accurate depth and atmospheric light estimation, both of which, however, are still open problems on their own. Therefore, the synthesized fog still suffers from artifacts. On the other hand, synthetic foggy im-



(a) TIT (b) AIT

Figure 1: Traditional image translation (TIT) vs. analogical image translation (AIT). Given images of two styles in the source domain $\mathcal{A}$ and $\mathcal{A}'$ with images of the first style $\mathcal{B}$ on the target domain, traditional methods can only translate between seen styles $\mathcal{A}, \mathcal{A}'$ and $\mathcal{B}$. The proposed analogical image translation is able to translate $\mathcal{B}$ to $\mathcal{B}'$, such that $\mathcal{A} : \mathcal{A}' :: \mathcal{B} : \mathcal{B}'$, without requiring any sample from $\mathcal{B}'$.

ages can be generated easily in virtual environments (Gaidon et al. 2016). This motivates the development of our AIT method which learns from the abundant synthetic clear-weather and synthetic foggy images to perform an analogical image translation from real clear-weather images to 'real' foggy images. AIT learns the correlation between synthetic clear-weather and synthetic foggy images, and then applies the learned knowledge to the real domain. We call such learned correlation the *gist* of translation and assume it transferable across domains. Since the proposed method uses analogy in the GAN setup, we call it AnalogicalGAN.

AnalogicalGAN achieves zero-shot translation ability by coupling a supervised training scheme in the synthetic domain, a cycle consistency strategy in the real domain, and an adversarial training scheme between the two domains. More specifically, in the synthetic domain, the *gist* of translation is learned in the supervised manner with the accessible paired clear-weather and foggy images. Then, this translation *gist* is transferred to the real domain through an adversarial learning scheme. In the real domain, the learning is further supervised through a cycle consistency scheme. The pipeline of AnalogicalGAN is shown in Fig. 2. Our extensive experiments demonstrate the superiority of AnalogicalGAN over the standard zero-shot image translation methods, when tested for fog generation. The quality of our fogy real images is also validated by the state-of-the-art performance on downstreamed semantic foggy scene understanding task.

Related Works

Image-to-Image Translation. Image translation methods have been developed to convert images of one given style to another given style with remarkable success in the last years (Zhu et al. 2017; Huang et al. 2018; Liu et al. 2019). Image translation is also becoming a standard step for domain adaptation (Tsai et al. 2018; Lian et al. 2019; Tsai et al. 2019; Zou et al. 2019; Vu et al. 2019; Hoffman et al. 2016; Chen, Li, and Van Gool 2018) – synthetic images are first translated to 'real' on which the downstream tasks such as segmentation and detection are then conducted (Hoffman et al. 2018; Chen et al. 2019; Li, Yuan, and Vasconcelos 2019; Gong et al. 2019; Dundar et al. 2018). The standard image translation frameworks (Zhu et al. 2017; Isola et al. 2017; Huang et al. 2018; Liu, Breuel, and Kautz 2017) require the availability of images of both styles involved in translation. To facilitate the translation to an unseen style, the proposed AIT exploit the concept of analogy from synthetic images, followed by its application on real images.

Image Analogy. The image analogy works (Hertzmann et al. 2001; Liao et al. 2017; Cheng, Vishwanathan, and Zhang 2008; Chen, Xu, and Jia 2020) aim to find $\mathcal{B}'$ related to $\mathcal{B}$ in the same way as $\mathcal{A}'$ relates to $\mathcal{A}$. Even though the purpose of image analogy is similar to that of our AIT, traditional image analogy works (Hertzmann et al. 2001; Liao et al. 2017; Cheng, Vishwanathan, and Zhang 2008) only apply coarse-to-fine filters to reduce the perceptual similarity distance, such as luminance feature (Hertzmann et al. 2001) and VGG feature (Liao et al. 2017) distances, between source and target domains. They do not disentangle the *gist* and have no knowledge transfer between the source and target domains. These works limit themselves, by design, to low-level applications such as the super resolution, artistic filters, and texture transfer (Hertzmann et al. 2001). In contrast, the high-level task of image translation in the concurrent work (Chen, Xu, and Jia 2020) combines GANs with analogical perceptual loss. Although the used perceptual loss is found to be effective for attribute manipulation, the same loss may not be sufficient for other image translation tasks (Huang et al. 2018). Furthermore, (Chen, Xu, and Jia 2020) does not exploit the concept of *gist*. Our work demonstrates that the *gist* can be effectively used for the task of analogical image translation.

Semantic Foggy Scene Understanding. Our work is also related to methods for semantic foggy scene understanding (SFSU). The SFSU aims to improve the performance of semantic scene understanding under foggy condition (Sakaridis, Dai, and Van Gool 2018; Dai et al. 2019; Hahner et al. 2019; Erkent and Laugier 2020; Tarel et al. 2010). Due to the difficulty of gathering and labeling large-scale foggy image dataset, some works (Sakaridis, Dai, and Van Gool 2018; Dai et al. 2019) synthesize fog by applying a physical model to the real clear weather images from the Cityscapes (Cordts et al. 2016), resulting the Foggy Cityscapes dataset. While yielding improved results, these methods require accurate depth and atmospheric light estimation. Any failure of these two tasks directly implies the failure of fog synthesis. The proposed AIT does not require to estimate atmospheric light, and uses depth only as an auxiliary information. Instead, AIT learns the necessary *gist* for translation from synthetic examples.

Unsupervised Domain Adaptation. AIT also shares some similarity with the unsupervised domain adaptation (UDA) works. UDA has been extensively studied in the past years, mainly for classification (Ganin and Lempitsky 2015; Long et al. 2018; Tzeng et al. 2017), semantic segmentation (Chen, Li, and Van Gool 2018; Tsai et al. 2019; Dai et al. 2019; Li, Yuan, and Vasconcelos 2019) and object detection (Chen et al. 2018; Xie et al. 2019; Zhu et al. 2019). Given a set of images and annotation pairs from the source domain, along with only the images from target, the goal is to learn a model that performs well also in the target domain. Our AIT shares the same spirit, in regard to transferring the

learned model from the source to target domain, without using annotations (images of desired styles) from the target domain. While previous UDA works only focus on the understanding tasks such as classification, object detection and segmentation, our work pays attention to the totally different task, *i.e.* image-to-image translation.

Analogical Image Translation

Problem Statement

In the image translation problem, we are given a source domain $\mathcal{S}$ and a target domain $\mathcal{T}$, which consist of the samples $\mathbf{x}^s \in \mathcal{S}$ and $\mathbf{x}^t \in \mathcal{T}$, respectively. The goal of traditional image translation is to transfer image samples $\mathbf{x}^s$ and $\mathbf{x}^t$ between domain $\mathcal{S}$ and domain $\mathcal{T}$. In our work, we propose analogical image translation (AIT), where the source domain $\mathcal{S}$ and the target domain $\mathcal{T}$ cover two styles $\mathcal{A}, \mathcal{A}'$ and $\mathcal{B}, \mathcal{B}'$, respectively. But during training and testing, there are only samples $\mathbf{x}^a \in \mathcal{A}$, $\mathbf{x}^{a'} \in \mathcal{A}'$ and $\mathbf{x}^b \in \mathcal{B}$ available. AIT aims at learning from available samples $\mathbf{x}^a, \mathbf{x}^{a\prime}$ to translate $\mathbf{x}^b$ to the unseen samples $\mathbf{x}^{b'}$, such that $\mathbf{x}^a : \mathbf{x}^{a'} :: \mathbf{x}^b : \mathbf{x}^{b'}$. The data distributions are denoted as $\mathbf{x}^a \sim P_A, \mathbf{x}^{a'} \sim P_{A'}, \mathbf{x}^b \sim P_B$ and $\mathbf{x}^{b'} \sim P_{B'}$. Our objective in this work is to learn the mapping $G_{BB'} : \mathcal{B} \rightarrow \mathcal{B}'$ conditioned on the mapping $G_{AA'} : \mathcal{A} \rightarrow \mathcal{A}'$. Note that, unlike our objective, the traditional methods (Zhu et al. 2017; Hoffman et al. 2018; Huang et al. 2018; Dundar et al. 2018) only focus on learning the mapping $G_{ST} : \mathcal{S} \rightarrow \mathcal{T}$.

AnalogicalGAN Model

In this section, we present our AnalogicalGAN for the analogical image translation problem. The key idea of AnalogicalGAN is to disentangle the translation *gist* in the source domain, transfer the *gist* to the target domain, and make the *gist* compatible with the target domain. In our work, the *gist* is measured with the alignment map $\mathcal{M}$ and the residual map $\mathcal{N}$, formally denoted as $\{\mathcal{M}, \mathcal{N}\}$. Taking the translation direction into account, the $\{\mathcal{M}, \mathcal{N}\}$ can be further expressed in detail as $\mathcal{M} = \{\mathcal{M}_{AA'}, \mathcal{M}_{A'A}, \mathcal{M}_{BB'}, \mathcal{M}_{B'B}\}, \mathcal{N} = \{\mathcal{N}_{AA'}, \mathcal{N}_{A'A}, \mathcal{N}_{BB'}, \mathcal{N}_{B'B}\}$. Moreover, the *gist* is assumed to be invariant to the source domain and the target domain. Then the *gist* can be defined implicitly as:

$$\mathcal{A}' = \mathcal{A} \odot \mathcal{M}_{AA'} + \mathcal{N}_{AA'}, \tag{1}$$

$$\mathcal{B}' = \mathcal{B} \odot \mathcal{M}_{BB\prime} + \mathcal{N}_{BB'}, \tag{2}$$

$$\mathcal{A} = \mathcal{A}' \odot \mathcal{M}_{A'A} + \mathcal{N}_{A'A}, \tag{3}$$

$$\mathcal{B} = \mathcal{B}' \odot \mathcal{M}_{B\prime B} + \mathcal{N}_{B'B}, \tag{4}$$

where $\odot$ denotes the element-wise multiplication. On this basis, as shown in Fig. 2, taking the direction of first style to second style for example, *i.e.* $\mathcal{A} \rightarrow \mathcal{A}'$, $\mathcal{B} \rightarrow \mathcal{B}'$, our framework consists of three main components: the supervised module, the adversarial module and the cycle consistent module. Firstly, on the source domain, due to the paired samples from $\mathcal{A}$ and $\mathcal{A}\prime$ available, the *gist*, $\mathcal{M}_{AA'}, \mathcal{N}_{AA'}$, is disentangled in the supervised way according to the Eq. (1), which forms the supervised module. Secondly, in the adversarial module, based on the domain invariant assumption of the *gist*, the *gist* on the source domain, $\mathcal{M}_{AA'}, \mathcal{N}_{AA'}$

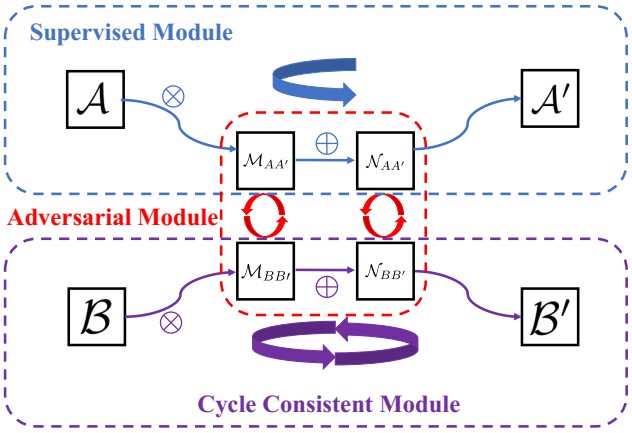


Figure 2: AnalogicalGAN overview. The AnalogicalGAN consists of three modules: the supervised, the adversarial, and the cycle-consistent. The supervised module disentangles the *gist*, $\mathcal{M}_{AA'}, \mathcal{N}_{AA'}$, using the supervised learning. The adversarial module transfers the learned *gist* from source domain, $\mathcal{M}_{AA'}, \mathcal{N}_{AA'}$, to the target domain, $\mathcal{M}_{BB'}, \mathcal{N}_{BB'}$. The cycle consistent module ensures that the transferred *gist* is compatible with the target domain.

, is transferred to the target domain, $\mathcal{M}_{BB'}, \mathcal{N}_{BB'}$, through the adversarial learning. Thirdly, on the target domain, due to the unavailability of the second style $\mathcal{B}'$, the *gist*, $\mathcal{M}_{BB'}, \mathcal{N}_{BB'}$ is retained to be compatible with the target domain through the cycle consistency, constructing the cycle consistent module. The other direction from the second style to the first style, $\mathcal{A}' \rightarrow \mathcal{A}, \mathcal{B}' \rightarrow \mathcal{B}$, acts in the same way. Next, the different modules and corresponding loss function are introduced in detail.

Supervised Module. The supervised module is used to disentangle the *gist*, $\mathcal{M}, \mathcal{N}$, from the source domain. Given the paired sample $\mathbf{x}^a \in \mathcal{A}$ and $\mathbf{x}^{a'} \in \mathcal{A}'$ on the source domain $\mathcal{S}$, the translation between $\mathcal{A}$ and $\mathcal{A}'$ can be trained in the supervised way, by substituting in Eq.(1), written as,

$$\mathcal{L}_{sup} = \mathbb{E}_{\mathbf{x}^a \sim P_A}\left[\|\mathbf{x}^a \odot \mathbf{m}^{aa'} + \mathbf{n}^{aa'} - \mathbf{x}^{a'}\|_1\right] \tag{5}$$

$$+\mathbb{E}_{\mathbf{x}^{a'} \sim P_{A'}}\left[\|\mathbf{x}^{a'} \odot \mathbf{m}^{a'a} + \mathbf{n}^{a'a} - \mathbf{x}^{a}\|_1\right], \tag{6}$$

where $(\mathbf{m}^{aa'}, \mathbf{n}^{aa'}) = G_{AA'}(\mathbf{x}^a)$ and $(\mathbf{m}^{a'a}, \mathbf{n}^{a'a}) = G_{A'A}(\mathbf{x}^{a'})$.

Adversarial Module. The adversarial module aims to transfer the *gist*, disentangled from the source domain, to the real domain. Specifically, taking the direction, $\mathcal{A} \rightarrow \mathcal{A}'$, $\mathcal{B} \rightarrow \mathcal{B}'$, for example, we introduce the discriminator D_I to distinguish the *gist* between the source domain, $\{\mathcal{M}_{AA'}, \mathcal{N}_{AA'}\}$, and the target domain, $\{\mathcal{M}_{BB'}, \mathcal{N}_{BB'}\}$. And the discriminator D_J acts in the same way in the inverse direction $\mathcal{A}' \rightarrow \mathcal{A}$, $\mathcal{B}' \rightarrow \mathcal{B}$. Then the adversarial loss of *gist* $\{\mathcal{M}, \mathcal{N}\}$ on $\mathcal{S}$ and $\mathcal{T}$ can be written as,

$$\begin{aligned}\mathcal{L}_{adv1}(G_{AA'}, G_{BB'}, D_I) \\ = \mathbb{E}_{\mathbf{x}^a \sim P_A}\left[\log(D_I(G_{AA'}(\mathbf{x}^a)))\right] \\ + \mathbb{E}_{\mathbf{x}^b \sim P_B}\left[\log(1 - D_I(G_{BB'}(\mathbf{x}^b)))\right].\end{aligned} \tag{7}$$

The similar adversarial loss $\mathcal{L}_{adv2}(G_{A'A}, G_{B'B}, D_J)$ is also defined for the direction $\mathcal{A}' \to \mathcal{A}$, $\mathcal{B}' \to \mathcal{B}$. Then the *gist* adversarial loss can be formulated as:

$$\mathcal{L}_{adv} = \mathcal{L}_{adv1} + \mathcal{L}_{adv2}. \tag{8}$$

In order to make the mapping $G_{BB'}$ conditional on $G_{AA'}$, the $G_{AA'}$ and $G_{BB'}$, $G_{A'A}$ and $G_{B'B}$ share all the parameters, respectively.

Cycle Consistent Module. The cycle consistent module is utilized to make the *gist* compatible with the target domain, *i.e.* preserve the target domain feature of the translated *gist*. Accordingly, the reconstruction loss is taken to recover $\mathbf{x}^b$ from the translated image $\mathbf{x}^{b'}$ through the inverse mapping $G_{B'B}$. Furthermore, in order to strengthen the recovery, another discriminator D_T is introduced to distinguish between the recovered $\mathbf{x}^b$ and the original $\mathbf{x}^b$. Then the image cycle consistency loss $\mathcal{L}_{cyc}$ consists of the reconstruction loss $\mathcal{L}_{rec}$ and the adversarial loss $\mathcal{L}_{adv}(G_{BB'}, G_{B'B}, D_T)$, by substituing in Eq. (2), given by:

$$\begin{aligned}
\mathcal{L}_{cyc} = &\quad \mathcal{L}_{rec} + \mathcal{L}_{adv}(G_{BB'}, G_{B'B}, D_T) && (9)\\
\mathcal{L}_{rec} = &\quad \mathbb{E}_{\mathbf{x}^b \sim P_B}\left[\|\mathbf{m}^{b'b} \odot (\mathbf{x}^{b'}) + \mathbf{n}^{b'b} - \mathbf{x}^b\|_1\right] && (10)\\
\mathcal{L}_{adv}&(G_{BB'}, G_{B'B}, D_T)\\
= &\quad \mathbb{E}_{\mathbf{x}^b \sim P_B}\left[\log(1 - D_T(\mathbf{m}^{b'b} \odot \mathbf{x}^{b'} + \mathbf{n}^{b'b}))\right] && (11)\\
+ &\quad \mathbb{E}_{\mathbf{x}^b \sim P_B}\left[\log(D_T(\mathbf{x}^b))\right],
\end{aligned}$$

where $(\mathbf{m}^{bb'}, \mathbf{n}^{bb'}) = G_{BB'}(\mathbf{x}^b)$, $(\mathbf{m}^{b'b}, \mathbf{n}^{b'b}) = G_{B'B}(\mathbf{x}^{b'})$ and $\mathbf{x}^{b'} = \mathbf{m}^{bb'} \odot \mathbf{x}^b + \mathbf{n}^{bb'}$.

Auxiliary Module. Besides the three main modules, the auxiliary module is added to assist the analogical image translation process and introduce the auxiliary information. From (Huang et al. 2018) and (Johnson, Alahi, and Fei-Fei 2016), the perceptual loss calculates the VGG feature distance $\Phi(\cdot)$ (Simonyan and Zisserman 2014) between the translated image and the reference image, and is proven to be able to assist the image translation process. Generalizing the perceptual loss to analogical image translation, the perceptual loss is given in the analogical way, formulated as,

$$\begin{aligned}
\mathbf{e}^S =& \Phi(\mathbf{x}^{a'}) - \Phi(\mathbf{x}^a) && (12)\\
\mathbf{e}^T =& \Phi(\mathbf{x}^{b'}) - \Phi(\mathbf{x}^b) && (13)\\
\mathcal{L}_{percep} =& \ \mathbb{E}_{\mathbf{x}^b \sim P_B}\left[\|\mathbf{e}^S - \mathbf{e}^T\|_1\right], && (14)
\end{aligned}$$

where $(\mathbf{m}^{bb'}, \mathbf{n}^{bb'}) = G_{BB'}(\mathbf{x}^b)$ and $\mathbf{x}^{b'} = \mathbf{m}^{bb'} \odot \mathbf{x}^b + \mathbf{n}^{bb'}$. Meanwhile, in terms of specific setting such as the analogical foggy image translation, the corresponding auxiliary information to fog effects, such as depth information (Fattal 2008; Sakaridis, Dai, and Van Gool 2018; Dai et al. 2019), can also be leveraged. By introducing the mapping $G_{IH} : \mathcal{A} \to \mathcal{H}_S, \mathcal{B} \to \mathcal{H}_T$ and $G_{JH} : \mathcal{A}' \to \mathcal{H}_S, \mathcal{B}' \to \mathcal{H}_T$, where $\mathcal{H}_S$ and $\mathcal{H}_T$ denote the depth domain corresponding to $\mathcal{S}$ and $\mathcal{T}$, composed of depth map $\mathbf{d}^S$ and $\mathbf{d}^T$, respectively. The auxiliary depth loss is given by,

$$\begin{aligned}
\mathcal{L}_{dep} = &\quad \mathbb{E}_{\mathbf{x}^a \sim P_A}\left[\|G_{IH}(\mathbf{x}^a) - \mathbf{d}^S\|_1\right] && (15)\\
+ &\quad \mathbb{E}_{\mathbf{x}^{a'} \sim P_{A'}}\left[\|G_{JH}(\mathbf{x}^{a'}) - \mathbf{d}^S\|_1\right]\\
+ &\quad \mathbb{E}_{\mathbf{x}^b \sim P_B}\left[\|G_{IH}(\mathbf{x}^b) - \mathbf{d}^T\|_1\right]\\
+ &\quad \mathbb{E}_{\mathbf{x}^{b'} \sim P_{B'}}\left[\|G_{JH}(\mathbf{x}^{b'}) - \mathbf{d}^T\|_1\right].
\end{aligned}$$

By sharing the network parameters between G_{IH}, $G_{AA'}$, and $G_{BB'}$, G_{JH}, $G_{A'A}$ and $G_{B'B}$ respectively, the depth information is implicitly encoded into our analogical translation process.

Full Objective. Integrating the losses defined above, our full objective for AnalogicalGAN can be defined as:

$$\mathcal{L} = \mathcal{L}_{adv} + \lambda_1 \mathcal{L}_{sup} + \lambda_2 \mathcal{L}_{cyc} + \lambda_3 \mathcal{L}_{dep} + \lambda_4 \mathcal{L}_{percep}, \tag{16}$$

where λ_1, λ_2, λ_3 and λ_4 are hyper-parameters used to balance different parts of training loss. Following the general manner for training the adversarial model, the full objective is trained in the minimax way, *i.e.* minimize the objective for generator while maximizing the objective for discriminator.

Domain Interpolation. Benefiting from the disentangled *gist*, our AnalogicalGAN is able to generate the intermediate domain between $\mathcal{B}$ and $\mathcal{B}'$ during testing stage. Following (Gong et al. 2019), the variable $z \in [0, 1]$ is used to measure the domainness. The intermediate domain between $\mathcal{B}$ and $\mathcal{B}'$ are denoted as $\mathcal{I}_B^{(z)}$. When $z = 0$, the intermediate domain $\mathcal{I}_B^{(z)}$ are identical to $\mathcal{B}$; and when $z = 1$, it is identical to $\mathcal{B}'$. In order to generated the intermediate domain, it is assumed that the *gist* between $\mathcal{B}$ and $\mathcal{B}'$ is linear. On the basis of the linear assumption and Eq. (2), the intermediate domain can be written as,

$$\mathcal{I}_B^{(z)} = \mathcal{B} \odot ((\mathcal{M}_{BB'} - 1) \times z + 1) + \mathcal{N}_{BB'} \times z. \tag{17}$$

Experiments

In this section, we evaluate our AnalogicalGAN for fog generation task. As aforementioned, our method consists of two domains: a source domain $\mathcal{S}$ and a target domain $\mathcal{T}$. On $\mathcal{S}$ and $\mathcal{T}$, there are two styles $\mathcal{A}$ and $\mathcal{A}'$, $\mathcal{B}$ and $\mathcal{B}'$ defined, respectively. Because training data for $\mathcal{B}'$ is unavailable, existing image translation methods can only be trained for $\mathcal{A}'$ and $\mathcal{B}$, which does not serve the exact purpose – generating data in $\mathcal{B}'$. Training standard translation methods on $\mathcal{A}'$ and $\mathcal{B}$, nevertheless, can be taken as baseline methods. In our experiments, we instantiate $\mathcal{S}$, $\mathcal{T}$, $\mathcal{A}$, $\mathcal{A}'$, $\mathcal{B}$ and $\mathcal{B}'$ as follows: *synthetic* as $\mathcal{S}$, *real* as $\mathcal{T}$, *synthetic, clear weather* as $\mathcal{A}$, *synthetic, foggy weather* as $\mathcal{A}'$, *real, clear weather* as $\mathcal{B}$, and *real, foggy weather* as $\mathcal{B}'$.

Analogical Image Translation

We conduct the analogical image translation experiments by regarding Virtual KITTI (Gaidon et al. 2016) as synthetic domain, while Cityscapes (Cordts et al. 2016) as real domain. The depth maps of Cityscapes are generated from pre-trained deep model developed in (Chang and Chen 2018).

Virtual KITTI. Virtual KITTI is a dataset consisting of 2136 photo-realistic synthetic clear weather images imitating the content and structure of KITTI dataset (Geiger et al.

2013), each of which has paired foggy weather image and corresponding depth map available.

Cityscapes. Cityscapes is a dataset covering 2975 real clear weather images taken from different European cities, which are densely labeled with 19 semantic categories.

We follow the training procedure, generators and discriminators structure as CycleGAN (Zhu et al. 2017). The Adam optimizer (Kingma and Ba 2015) is adopted, the learning rate is fixed to 0.0002 and the batch size is set as 1. The image is resized to 512×256. The weight of the *gist* adversarial loss is set as 3, the weight of cycle consistency adversarial loss is set as 1, and the weight of rest parts are 10. We implement our model with PyTorch (Paszke et al. 2017). More detailed network architecture and implementation are shown in Appendix due to space limitation.

Gist. In order to verify the necessity of *gist* for the AIT task, the state-of-the-art traditional image translation frameworks CyelGAN (Zhu et al. 2017), and recent domain adaptive image translation framework DAI2I (Chen, Xu, and Jia 2020) are taken as baseline methods. When applied to AIT task, the traditional frameworks do not disentangle the *gist* from the source domain $\mathcal{S}$. For example, CycleGAN is only able to translate between $\mathcal{S}$ ($\mathcal{A}$, $\mathcal{A}'$) and $\mathcal{T}$ ($\mathcal{B}$). The direct translation between $\mathcal{S}$ and $\mathcal{T}$, if performed, unavoidably translates the feature irrelevant to *gist*. Such translation causes the artifacts in the generated $\mathcal{B}'$. In the fog generation case, the translation from *real, clear weather* to *synthetic, clear weather* (similarly for *real clear* to *synthetic foggy* weather) introduces the *synthetic* style to the *real* domain. It causes the generated *real, foggy weather* images inherit abundant of *synthetic* artifacts, instead of only *foggy weather* style. Though auxiliary information such as depth can be encoded into the CycleGAN baseline, by network parameters sharing as done in Eq.(16), it still cannot resolve the *synthetic* artifacts problem. Fig. 3b shows the qualitative results of the CycleGAN baseline translating between $\mathcal{A}\prime$ and $\mathcal{B}$ while using the depth information. It is observed that the translated *real, foggy weather* image with CycleGAN baseline is highly affected by the translated *synthetic* style. In order to improve CycleGAN baseline for AIT task, we also trained the model to translate from $\mathcal{A}$ to $\mathcal{A}'$ while testing on $\mathcal{B}$. This result in the reduction of the *synthetic* effect. From the comparison between Fig. 3b and Fig. 5d, it can be seen that the CycleGAN baseline's performance is improved, when $\mathcal{T}$ is eliminated during training. Nevertheless, these results are not yet satisfactory.

One can also think of using multi-stage translation strategy. A typical case could follow $\mathcal{B} \rightarrow \mathcal{A} \rightarrow \mathcal{A}'$, This multi-stage translation is bound to have synthetic effect at the end, because the final domain, *i.e.* $\mathcal{A}'$, is still synthetic. In essence, DAI2I (Chen, Xu, and Jia 2020) combines the aforementioned multi-stage translation strategy with the analogical perceptual similarity measurement, whose results are shown in Fig. 3c. Though image analogy spirit through the perceptual similarity is adopted in DAI2I, the *gist* is still not exploited. Fig. 3c shows that DAI2I cannot deal with *synthetic* artifacts purely relying on the analogical perceptual similarity for fog generation, without exploiting *gist*. Contrastively, by explicitly disentangling *gist* from $\mathcal{S}$ and transferring to $\mathcal{T}$, our AnalogicalGAN eliminates the *synthetic* artifacts. The qualitative results in Fig. 3c show a clear distinction between DAI2I and AnalogicalGAN. We choose not to conduct further analysis of DAI2I results, as they are clearly inferior in qualitative measure. The artifacts on the reported images are consistent across test set (more images can be found in the appendix).

$\mathcal{L}_{adv}$			✓	✓	✓	✓	✓
$\mathcal{L}_{cyc}$		✓		✓	✓	✓	✓
$\mathcal{L}_{percep}$		✓	✓			✓	✓
$\mathcal{L}_{dep}$		✓	✓		✓		✓
mIoU	34.6	32.8	42.0	41.7	41.9	40.8	**42.3**

Table 1: Full model and ablations comparison for SFSU, tested on the Foggy Zurich dataset based on RefineNet with ResNet-101 backbone. The results are reported on mIoU over 19 classes. The best result is denoted in bold.

Figure 3: Qualitative translation results of CycleGAN encoding depth and DAI2I (Chen, Xu, and Jia 2020). (a) is *real, clear weather* image, while (b), (c) are translated *real, foggy weather* image with CycleGAN model encoding depth and DAI2I, respectively. Both of CycleGAN and DAI2I introduce high *synthetic* artifacts to the translated images.

Quantitative Results. In order to validate the effectiveness of our AnalogicalGAN for the AIT task, a user study on Amazon Mechanical Turk (AMT) is conducted to compare the translation results of our AnalogicalGAN with the state-of-the-art traditional image translation methods CycleGAN (Zhu et al. 2017) and MUNIT (Huang et al. 2018). Each individual task completed by the participants, referred to as Human Intelligence Task (HIT), comprises two image pairs to be compared: ours vs. CycleGAN and ours vs. MUNIT. In total, 100 HITs were used, each is completed by three annotators and the results are averaged. For each image pair, the users were asked to select the image that looks more like a real foggy image. In Table 3, the user study results are listed. From the table, one can see that users prefer our translation results compared to CycleGAN (61.0% v.s. 39.0%) and MUNIT (66.7% v.s. 33.3%).

Qualitative Results. Furthermore, we show the qualitative comparison in Fig. 5. From Fig. 5, it is observed that the standard image translation models CycleGAN (refer to Fig. 5d) and MUNIT (refer to Fig. 5(e)) suffer from inheriting synthetic features from the Virtual KITTI (refer to Fig. 5(b)) such as the color of the car, the lines on the road and the skin of the people. Besides, though the translated foggy part tends to be in gray, it loses the correct sense that fog changes

Virtual KITTI→ Cityscapes					Virtual KITTI→ Synscapes				
Fine-tuning	Testing				Fine-tuning	Testing			
	FZ		FD			FZ		FD	
	R	B	R	B		R	B	R	B
Cityscapes(Hahner et al. 2019)	34.6	16.1	44.3	27.2	Cityscapes(Hahner et al. 2019)	34.6	16.1	44.3	27.2
FC(Hahner et al. 2019)	36.9	25.0	46.1	30.3	FS(Hahner et al. 2019)	40.3	27.8	48.4	30.9
CycleGAN(Zhu et al. 2017)	40.5	27.1	47.7	30.0	CycleGAN(Zhu et al. 2017)	41.6	30.9	47.8	33.1
MUNIT(Huang et al. 2018)	39.1	26.0	**47.8**	30.5	MUNIT(Huang et al. 2018)	40.5	27.5	48.3	32.8
AC(ours)	**42.3**	**28.4**	47.5	**30.8**	AS(ours)	**41.8**	**31.5**	**49.8**	**34.2**
(a)					(b)				

Table 2: Results of semantic segmentation on the Foggy Zurich and Foggy Driving dataset. The reported results are pretrained on Cityscapes, fine-tuned on different simulated foggy images, and tested on Foggy Zurich (FZ) and Foggy Driving (FD) datasets. The columns represent different semantic segmentation architectures, RefineNet (R) with ResNet-101 backbone and BiseNet (B) with ResNet-18 backbone. The results are reported on mIoU over 19 categories. The best results are denoted in bold. "FC", "FS", "AC", "AS", "FD", "FZ" represent "Foggy Cityscapes", "Foggy Synscapes", "AnalogicalGAN Cityscapes", "AnalogicalGAN Synscapes", "Foggy Driving", "Foggy Zurich", respectively.

	CycleGAN/Ours	MUNIT/Ours
user preference	39.0%/61.0%	33.3%/66.7%

Table 3: User study results for fog generation. It is observed that more users prefer the translation results of our AnalogicalGAN compared to that of CycleGAN and MUNIT.

with depth. In contrast, our AnalogicalGAN, the analogical image translation framework, preserves the real feature of the objects in the scene, generates realistic foggy images and yields the right sense that fog changes with the depth of the scene as shown in Fig. 5(f).

Ablation Study. Fig. 4 gives an qualitative ablation study of each module in our AnalogicalGAN. More qualitative ablation study results are put into Appendix due to space limitation.

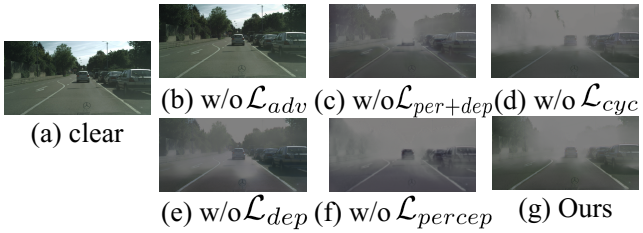


Figure 4: Qualitative ablation study of AnalogicalGAN for fog generation. It is observed that each module is effective for the analogical image translation (AIT) task.

Semantic Foggy Scene Understanding

Experiments Setup In this section, we validate the usefulness of our translated images for the downstream task semantic foggy scene understanding. Specifically, following the paradigm in (Sakaridis, Dai, and Van Gool 2018; Hahner et al. 2019), the pretrained semantic segmentation model on the real clear weather images, Cityscapes, is fine-tuned on the synthesized foggy images. Then the fine-tuned model is tested on two real foggy image datasets: Foggy Zurich(Dai et al. 2019) and Foggy Driving (Sakaridis, Dai, and Van Gool 2018). We compare the semantic foggy scene understanding performance of our AnalogicalGAN translation results with the state-of-the-art physics-based foggy image synthesis results, Foggy Cityscapes(Sakaridis, Dai, and Van Gool 2018), and the translation results of the traditional image translation methods CycleGAN and MUNIT. In addition to the setting *Virtual KITTI to Cityscapes*, we further evaluate all methods in another setting *Virtual KITTI to Synscapes*. The performance of foggy scene understanding of all methods are reported for both of the two settings.

Fine-tuning	Foggy Zurich		Foggy Driving	
	R	B	R	B
FC+FS(Hahner et al. 2019)	41.4	30.9	**50.7**	35.2
AC+FS (ours)	**43.8**	**32.9**	50.3	**39.9**

Table 4: Results of semantic segmentation on the Foggy Zurich (FZ) and Foggy Driving (FD) dataset. The reported results are pretrained on Cityscapes, fine-tuned on Foggy Cityscape (FC)/AnalogicalGAN Cityscapes(AC) and Foggy Synscape (FS), and tested on FZ and FD datasets. The columns represent different semantic segmentation architectures, RefineNet (R) with ResNet-101 backbone and BiseNet (B) with ResNet-18 backbone. The results are reported on mIoU over 19 categories, and best results are bold.

Synscapes is a synthetic dataset consisting of 25,000 clear weather images imitating the content and structure of Cityscapes dataset. Pixel-wise ground-truth semantic labels and depth maps are given in the dataest.

Foggy Zurich consists of 3,808 foggy scene images taken from Zurich City, 40 of which are densely labeled. We use them as test data in our experiment.

Foggy Driving is a dataset containing 101 coarsely annotated real foggy images, collected in various areas of Zurich and from the Internet.

As shown in (Dai et al. 2019), the fog density of the synthesized foggy image highly affects the semantic foggy

(a) Clear VKITTI (b) Foggy VKITTI (c) Clear Cityscapes (d) CycleGAN (e) MUNIT (f) Ours

Figure 5: Comparison of the analogical translation results of our AnalogicalGAN (column (f)) with the traditional image translation methods (column (d) and column (e)). The column (a), column (b) and column (c) shows the synthetic clear weather image (Clear Virtual KITTI), the synthetic foggy weather image (Foggy Virtual KITTI) and the real clear weather image (Cityscapes), respectively. The analogical translation is described as, column (a) : column (b) :: column (c) : column (d), column (e), column (f).

scene understanding performance. Our AnalogicalGAN can control the density of the synthesized fog via the domainness variable z. In order to generate the foggy image with the appropriate fog density, during testing stage, the domainness variable z is set to 0.88 and 0.9 for Cityscapes and Synscapes, respectively. For semantic segmentation, we follow the paradigm and fine-tuning details in (Sakaridis, Dai, and Van Gool 2018) and (Hahner et al. 2019). The RefineNet (Lin et al. 2017) with ResNet-101 backbone (He et al. 2016) and the BiseNet (Yu et al. 2018) with ResNet-18 backbone (He et al. 2016) are utilized as the segmentation networks.

Experiments Results The results of semantic foggy scene understanding based on the synthesized foggy images from Cityscapes and Synscapes are shown in Table 2a and Table 2b, respectively. In Table 2a and Table 2b, while using Cityscapes and Synscapes as real clear weather images, it is shown that our AnalogicalGAN outperforms the physics-based foggy image synthesis methods "Foggy Cityscapes" and "Foggy Synscapes". The improvement is consistent on both Foggy Zurich and Foggy Driving, and with RefineNet and with BiseNet segmentatin networks. When compared to the traditional image translation methods, our "AnalogicalGAN" outperforms both "CycleGAN" and "MUNIT" on both test sets and for both segmentation networks, except for one case (when utilizing the RefineNet and testing on Foggy Driving) in which our method reaches comparable performance with MUNIT (47.5% v.s. 47.8%).

Moreover, following (Hahner et al. 2019), by mixing the "Foggy Synscapes" with "AnalogicalGAN Cityscapes", *i.e.* Cityscapes translated with "AnalogicalGAN" model, the performance can be further improved. From Table 4, it is shown that the mixture of "AnalogicalGAN Cityscapes" and "Foggy Synscapes" improves the performance of the state-of-the-art methods, mixture of "Foggy Citysacpes" and "Foggy Synscapes" by 2.4% and 2.0% on Foggy Zurich with RefineNet and BiseNet, while improving by 4.7% on Foggy Driving with BiseNet and reaching comparable performance, 50.3% v.s. 50.7%, on Foggy Driving with RefineNet. The semantic foggy scene understanding performance and comparison demonstrate the effectiveness of our AnalogicalGAN for synthesizing fog effects to real images. The results also shows the advantage of our proposed method over the physics-based fog synthesis methods and the traditional image translation methods. More detailed results on each classes are listed in the Appendix due to the space limitation.

Ablation Study. In Table 1, we compare our model with the ablations of the full objective for the semantic foggy scene understanding, *i.e.* quantitative ablation study results. It is shown that each module of our AnalogicalGAN contributes to the semantic foggy scene understanding.

Conclusion

In this work, we have presented AnalogicalGAN, a novel analogical image translation (AIT) framework. Different from the traditional image translation, analogical image translation is able to achieve zero-shot image translation capability via analogy. Compared with previous image analogy works, our AnalogicalGAN explicitly disentangles *gist* and transfers *gist*, which is proven to be necessary and beneficial for the AIT task. Applying our AnalogicalGAN to the fog generation task, the realistic fog effects is synthesized into real clear-weather images, even though no real foggy image is ever seen. Further experiments prove the effectiveness of our AnalogicalGAN. While some choices in AnalogicalGAN are made specifically for fog generation, the method itself has the potential to be used for other AIT tasks.

Acknowledgements

This research has received funding from the EU Horizon 2020 research and innovation programme under grant agreement No. 820434. This work is funded by Toyota Motor Europe via the research project TRACE Zurich. This work is also partially supported by the Major Project for New Generation of AI under Grant No. 2018AAA0100400.

Ethics Statement

In this paper, we propose the "AnalogicalGAN" model, a kind of analogical image translation framework. It can be seen as the zero-shot generalization of existing image-to-image translation framework.

The analogical image translation framework has the potential to highly reduce the gathering and labeling difficulty of the data. Benefiting from the transferred data scale and diversity, the deep model is expected to be more robust, reliable and effective under different even extreme conditions, which is able to promote and accelerate the launch of deep-based system such as the medical computer-assisted system and autonomous driving system.

The easy availability of the transferred labeled data and the launch of the more reliable and effective deep-based systems likely have complex social impacts. (i) On one hand, transferred labeled data will save much cost on the data gathering and labeling and avoid the wasteful duplication of labor. More and more deep-based artificial intelligent systems will become part of the people's life, bringing convenience, wealth and prosperity. (ii) On the other hand, the transferred labeled data might induce the unemployment for the people who are engaged in gathering and labeling the dataset. Meanwhile, the launch of artificial intelligent systems may also cause the job loss. Besides, another concern is that the techniques for synthesizing the image is possible to be used for the illegal purpose of forgery and deception.

We would encourage further work on the detection of the forgery and deception of the image even though the detection will become harder and harder as the image synthesis techniques develop. From the view of long-term development, in order to mitigate the risks of image synthesis, more regulations and guidance on tracking and stopping the harmful and dangerous synthesized images should be made.

References

Blum, H.; Sarlin, P.-E.; Nieto, J.; Siegwart, R.; and Cadena, C. 2019. Fishyscapes: A benchmark for safe semantic segmentation in autonomous driving. In *ICCV Workshops*.

Chang, J.-R.; and Chen, Y.-S. 2018. Pyramid stereo matching network. In *CVPR*.

Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; and Yuille, A. L. 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence (TPAMI)* 40(4): 834–848.

Chen, Y.; Li, W.; Chen, X.; and Gool, L. V. 2019. Learning semantic segmentation from synthetic data: A geometrically guided input-output adaptation approach. In *CVPR*.

Chen, Y.; Li, W.; Sakaridis, C.; Dai, D.; and Van Gool, L. 2018. Domain adaptive faster r-cnn for object detection in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Chen, Y.; Li, W.; and Van Gool, L. 2018. Road: Reality oriented adaptation for semantic segmentation of urban scenes. In *CVPR*.

Chen, Y.-C.; Xu, X.; and Jia, J. 2020. Domain Adaptive Image-to-image Translation. In *CVPR*.

Cheng, L.; Vishwanathan, S. N.; and Zhang, X. 2008. Consistent image analogies using semi-supervised learning. In *CVPR*.

Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; and Schiele, B. 2016. The cityscapes dataset for semantic urban scene understanding. In *CVPR*.

Dai, D.; Sakaridis, C.; Hecker, S.; and Van Gool, L. 2019. Curriculum model adaptation with synthetic and real data for semantic foggy scene understanding. *International Journal of Computer Vision* 1–23.

Dundar, A.; Liu, M.-Y.; Wang, T.-C.; Zedlewski, J.; and Kautz, J. 2018. Domain Stylization: A Strong, Simple Baseline for Synthetic to Real Image Domain Adaptation. *arXiv preprint arXiv:1807.09384* .

Erkent, Ö.; and Laugier, C. 2020. Semantic Segmentation with Unsupervised Domain Adaptation Under Varying Weather Conditions for Autonomous Vehicles. *IEEE Robotics and Automation Letters* 5(2): 3580–3587.

Fattal, R. 2008. Single image dehazing. *ACM transactions on graphics (TOG)* 27(3): 1–9.

Gaidon, A.; Wang, Q.; Cabon, Y.; and Vig, E. 2016. Virtual Worlds as Proxy for Multi-Object Tracking Analysis. In *CVPR*.

Ganin, Y.; and Lempitsky, V. 2015. Unsupervised domain adaptation by backpropagation. In *ICML*.

Geiger, A.; Lenz, P.; Stiller, C.; and Urtasun, R. 2013. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research* 32(11): 1231–1237.

Gong, R.; Li, W.; Chen, Y.; and Gool, L. V. 2019. DLOW: Domain flow for adaptation and generalization. In *CVPR*.

Hahner, M.; Dai, D.; Sakaridis, C.; Zaech, J.-N.; and Van Gool, L. 2019. Semantic Understanding of Foggy Scenes with Purely Synthetic Data. In *ITSC*.

Halder, S. S.; Lalonde, J.-F.; and Charette, R. d. 2019. Physics-Based Rendering for Improving Robustness to Rain. In *ICCV*.

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

Hertzmann, A.; Jacobs, C. E.; Oliver, N.; Curless, B.; and Salesin, D. H. 2001. Image Analogies. In *Annual Conference on Computer Graphics and Interactive Techniques*.

Hoffman, J.; Tzeng, E.; Park, T.; Zhu, J.-Y.; Isola, P.; Saenko, K.; Efros, A. A.; and Darrell, T. 2018. CyCADA: Cycle Consistent Adversarial Domain Adaptation. In *ICML*.

Hoffman, J.; Wang, D.; Yu, F.; and Darrell, T. 2016. Fcns in the wild: Pixel-level adversarial and constraint-based adaptation. *arXiv preprint arXiv:1612.02649* .

Huang, X.; Liu, M.-Y.; Belongie, S.; and Kautz, J. 2018. Multimodal Unsupervised Image-to-image Translation. In *ECCV*.

Isola, P.; Zhu, J.-Y.; Zhou, T.; and Efros, A. A. 2017. Image-to-image translation with conditional adversarial networks. In *CVPR*.

Johnson, J.; Alahi, A.; and Fei-Fei, L. 2016. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*.

Kingma, D. P.; and Ba, J. 2015. Adam: A method for stochastic optimization. In *ICLR*.

Li, K.; Li, Y.; You, S.; and Barnes, N. 2017. Photo-realistic simulation of road scene for data-driven methods in bad weather. In *ICCV Workshops*.

Li, Y.; Yuan, L.; and Vasconcelos, N. 2019. Bidirectional learning for domain adaptation of semantic segmentation. In *CVPR*.

Lian, Q.; Lv, F.; Duan, L.; and Gong, B. 2019. Constructing Self-Motivated Pyramid Curriculums for Cross-Domain Semantic Segmentation: A Non-Adversarial Approach. In *ICCV*.

Liao, J.; Yao, Y.; Yuan, L.; Hua, G.; and Kang, S. B. 2017. Visual attribute transfer through deep image analogy. In *SIGGRAPH*.

Lin, G.; Milan, A.; Shen, C.; and Reid, I. 2017. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *CVPR*.

Liu, M.-Y.; Breuel, T.; and Kautz, J. 2017. Unsupervised image-to-image translation networks. In *NIPS*.

Liu, M.-Y.; Huang, X.; Mallya, A.; Karras, T.; Aila, T.; Lehtinen, J.; and Kautz, J. 2019. Few-Shot Unsupervised Image-to-Image Translation. In *The IEEE International Conference on Computer Vision (ICCV)*.

Long, M.; Cao, Z.; Wang, J.; and Jordan, M. I. 2018. Conditional adversarial domain adaptation. In *NIPS*.

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

Ren, W.; Liu, S.; Zhang, H.; Pan, J.; Cao, X.; and Yang, M.-H. 2016. Single image dehazing via multi-scale convolutional neural networks. In *ECCV*.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*.

Sakaridis, C.; Dai, D.; and Van Gool, L. 2018. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision (IJCV)* 126(9): 973–992.

Schunn, C. D.; and Dunbar, K. 1996. Priming, analogy, and awareness in complex reasoning. *Memory & Cognition* 24(3): 271–284.

Simonyan, K.; and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* .

Tarel, J.-P.; Hautiere, N.; Cord, A.; Gruyer, D.; and Halmaoui, H. 2010. Improved visibility of road scene images under heterogeneous fog. In *IEEE Intelligent Vehicles Symposium*.

Tsai, Y.-H.; Hung, W.-C.; Schulter, S.; Sohn, K.; Yang, M.-H.; and Chandraker, M. 2018. Learning to Adapt Structured Output Space for Semantic Segmentation. In *CVPR*.

Tsai, Y.-H.; Sohn, K.; Schulter, S.; and Chandraker, M. 2019. Domain adaptation for structured output via discriminative patch representations. In *CVPR*.

Tzeng, E.; Hoffman, J.; Saenko, K.; and Darrell, T. 2017. Adversarial discriminative domain adaptation. In *CVPR*.

Vu, T.-H.; Jain, H.; Bucher, M.; Cord, M.; and Pérez, P. 2019. ADVENT: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation. In *CVPR*.

Xie, R.; Yu, F.; Wang, J.; Wang, Y.; and Zhang, L. 2019. Multi-level Domain Adaptive learning for Cross-Domain Detection. In *IEEE International Conference on Computer Vision Workshops*.

Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; and Sang, N. 2018. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In *ECCV*.

Yu, F.; and Koltun, V. 2016. Multi-scale context aggregation by dilated convolutions. In *ICLR*.

Zhao, H.; Shi, J.; Qi, X.; Wang, X.; and Jia, J. 2017. Pyramid scene parsing network. In *CVPR*.

Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*.

Zhu, X.; Pang, J.; Yang, C.; Shi, J.; and Lin, D. 2019. Adapting Object Detectors via Selective Cross-Domain Alignment. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Zou, Y.; Yu, Z.; Liu, X.; Kumar, B. V.; and Wang, J. 2019. Confidence Regularized Self-Training. In *ICCV*.