





Bridging the Domain Gap towards Generalization in Automatic Colorization

Hyejin Lee¹, Daehee Kim^{1,2}, Daeun Lee³,
Jinkyu Kim^{4*}, and Jaekoo Lee^{1*}

¹ Department of Computer Science, Kookmin University

² Clova AI Research, NAVER Corp.

³ Department of Statistics, Korea University

⁴ Department of Computer Science and Engineering, Korea University

*Corresponding authors: jinkyukim@korea.ac.kr, jaekoo@kookmin.ac.kr

Abstract. We propose a novel automatic colorization technique that learns domain-invariance across multiple source domains and is able to leverage such invariance to colorize grayscale images in unseen target domains. This would be particularly useful for colorizing sketches, line arts, or line drawings, which are generally difficult to colorize due to a lack of data. To address this issue, we first apply existing domain generalization (DG) techniques, which, however, produce less compelling desaturated images due to the network’s over-emphasis on learning domain-invariant contents (or shapes). Thus, we propose a new domain generalizable colorization model, which consists of two modules: (i) a domain-invariant content-biased feature encoder and (ii) a source-domain-specific color generator. To mitigate the issue of insufficient source domain-specific color information in domain-invariant features, we propose a skip connection that can transfer content feature statistics via adaptive instance normalization. Our experiments with publicly available PACS and Office-Home DG benchmarks confirm that our model is indeed able to produce perceptually reasonable colorized images. Further, we conduct a user study where human evaluators are asked to (1) answer whether the generated image looks naturally colored and to (2) choose the best-generated images against alternatives. Our model significantly outperforms the alternatives, confirming the effectiveness of the proposed method. The code is available at <https://github.com/Lhyejin/DG-Colorization>.

Keywords: Automatic Colorization, Domain Generalization, Generative Adversarial Networks

1 Introduction

Recent successes in applying deep learning to computer vision tasks suggest that a ConvNet-based model can learn a fully automated data-driven colorization of grayscale images. This model can be trained to predict the a and b color channels in the CIE-*Lab* color space using semantic information and the surface texture of given grayscale image [1,6,15,43]. In practice, data-driven colorization is used

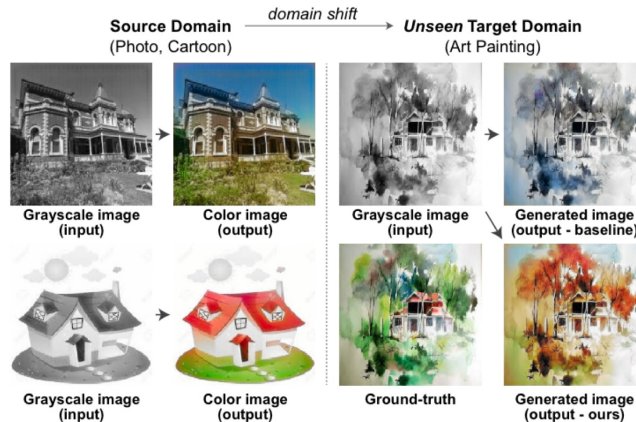


Fig. 1. We propose a data-driven fully automated colorization technique that can learn a mapping function from the grayscale image to color channels by regression onto continuous color space. With a novel domain generalization technique, our model can generalize well to unseen target domains, whereas existing colorization approaches and conventional DG techniques often fail to generate.

in applications such as coloring assistance of cartoonist or legacy photo restoration. However, current learning-based methods often fail to generalize colorizing performance in out-of-distribution. While successfully producing plausible and visually compelling colorization within the same known test domain, such models would not intrinsically generalize well to novel domains outside the training distribution because different domains involve different textures [11,18,30]. For example, a model trained with photo and cartoon images would fail to colorize art painting images (see Fig. 1). To address this issue, we first analyze the effect of domain shift in the colorization task and propose a novel domain generalization technique that enables a colorization model to generalize well to novel domains outside the training distribution. This model would be particularly useful for colorizing sketches, line arts, or line drawings, which are generally difficult to colorize due to a lack of data. Also, we would emphasize that, to our best knowledge, improving the model’s ability to generalize across multiple domains for colorization has not yet been explored.

In literature, various domain generalization techniques have been introduced to make models generalize well to out-of-training-distribution. These techniques primarily focused on generalizable object class classification models by learning the domain-invariances across multiple source domains so that the classifier can robustly leverage such invariances in unseen target domains [10,30,33,35]. Examples from target domains are not available during training, thus these approaches differ from domain adaptation, semi-supervised domain adaptation, and unsupervised domain generations.

In our experiment, applying existing DG techniques to the colorization models was not successful. We have found that these models tend to over-emphasize

generating domain-invariant contents and fail to leverage source domain-specific color information. Interestingly, such domain-specific information is particularly useful for colorization, as it can provide a domain-sensitive and essential prior for synthesizing colorized images of better quality.

To mitigate this issue, we propose the domain generalizable color generation network, which is divided into two components: (i) domain-invariant content-biased encoder and (ii) source domain-specific color generator. The former is trained to match content information across multiple domains to generalize well across domains by learning a domain-invariant texture (i.e., shapes), whereas the latter captures source domain-specific color information. Because the extracted features delivered through the skip connection between the encoder and decoder have little source domain-specific color information, it is necessary to adjust the color information to the source domain. Thus, we propose to transfer domain-invariant content feature statistics from encoder to decoder via skip connections followed by adaptive instance normalization (AdaIN) [14]. We observe such a connection significantly improves the quality of colored images.

We use two publicly available domain generalization benchmark datasets (PACS [23] and Office-Home [37]) to evaluate the effectiveness of the proposed method. We also analyze and compare with existing state-of-the-art domain generalization techniques, and we observe that our model generally outperforms these alternatives. In addition, we conduct a user study where human evaluators are asked to answer the following two questions: (i) Naturalness: “Do you think the provided image looks naturally colored” and (ii) Perceptual Realism: “Which of the images are the best”. Our work significantly outperforms others, which confirms the effectiveness of the proposed method. We summarize our contribution as follows:

- We propose a novel fully automated colorization method that can generalize well to unseen target domains.
- Our model utilizes domain-invariant contents-biased encoder and source domain-specific color generator where a skip connection is used to transfer domain-invariant content feature statistics between the encoder and the decoder for a better quality colorization.
- We effectively show the benefit of our proposed methods on two large domain generalization benchmark datasets: PACS and Office-home. Our proposed method quantitatively and qualitatively outperforms alternative colorization approaches and domain generalization techniques.
- We conduct a user study where human participants evaluate the quality in terms of naturalness and the perceptual realism. Our method significantly outperforms other alternative domain generational approaches.

2 Related Work

2.1 Image Colorization

Colorization algorithms use various approaches, which can be broadly categorized into the following three types: (i) scribble-based colorization techniques, (ii)

example-based colorization, and (iii) fully automatic colorization. Given an input grayscale image, the scribble-based colorization propagates scribbles (provided by the user) to the whole image [21]. Example-based colorization techniques, however, exploit user-provided [5,12,16,40], or automatically retrieved [7,27] reference images to match the luminance and texture information between a reference image and the input image (i.e., transferring color onto the input grayscale image from analogous regions of the reference image). These scribble-based and example-based approaches, though promising, depend primarily on user input, which can be time-consuming and expensive for achieving an acceptable result.

Recent work suggests that a data-driven, fully automated approach can successfully learn a mapping function from the lightness channel to color channels by regressing onto continuous color space or classifying them into quantized color values [3,6,15,17,31,34,38,43]. These approaches have developed similar systems, such as (i) convolutional neural networks, which are trained end-to-end to predict color channels of the image from the lightness channel using large-scale data, (ii) conditional GANs, which have a sharpening effect in the spectral dimension and make images more colorful, have recently become a key architectural component for colorization, (iii) Isola et al. [17] used a U-Net-based architecture [32] for the generator and a convolutional PatchGAN classifier [22] for the discriminator, and they got promising colorization results. Our method is also based on (iii).

Most of these fully automatic colorization techniques produce promising results in the same known test domains, but their generation quality tends to become sub-optimal in unseen different test domains. In previous work [34,38] because the pretrained ImageNet and COCO models were also used, domain shift occurred when training domains other than photo. Improving such ability to generalize across multiple domains has not yet been investigated (though important) in the community to the best of our knowledge. Here, we explore the effect of domain shift in the colorization task, and propose a fully automatic colorization model that can generalize well to novel domains outside the training distribution.

2.2 Domain Generalization

Domain generalization (DG) techniques focus on generating domain-invariant latent representations so that the model is to better generalize to the unseen target domains outside the training distribution. Vapnik et al. [36] introduces Empirical Risk Minimization (ERM) that minimizes the sum of errors across the domains of landmark work. Notable variants have been proposed to learn domain-invariant features by matching distributions across different domains. Ganin et al. [10] use an adversarial network for such distribution matching, while Li et al. [26] match the conditional distributions across the domains. Minimizing maximum mean discrepancy [25], transformed feature distribution distance [29], or covariances [35] is frequently used to optimize such a shared feature space. In this study, we also follow this workstream; however, we focus on the benefits of these techniques for the image colorization task, specifically how the model

can learn domain-invariant content information across different domains and generalize effectively to unknown target domains.

Inter-domain mixup [39,41,42] techniques are used on linearly interpolated examples from random pairs across domains to perform an ERM. JiGen [4] improves generalization using self-supervised clues obtained by solving a jigsaw puzzle as a secondary task. Meta-learning frameworks [24] are also investigated for DG to meta-learn how to generalize across domains by leveraging MAML [9]. Low-rank parameterization [23], style-agnostic network [30], and domain-specific aggregation modules [8] have recently been used to partition the model into domain-invariant and domain-variant components.

In stylized ImageNet [11] and Kim et al. [18], texture and content are defined respectively as a domain-specific feature and domain-invariant feature. They applied a random representation mixture to training data with AdaIN to make the extracted feature robust. Here, we also use such disentangled image representations (i.e., content vs. texture). However, we advocate learning not only a content-biased network (i.e., reducing domain-specific texture information), but also source domain-specific color information for the colorization task. As the latter can provide domain-sensitive content information about source domain-specific color features, which can serve as an essential prior to generate high-quality colorized images.

3 Method

3.1 Conditional GANs for Colorization

Our model is built upon the pix2pix architecture [17]. This model uses GANs in the conditional setting, which is suitable for image-to-image mappings as they can generate a corresponding image conditioned on an input image. As shown in Fig. 2, our model consists of two main components: (i) a generator $G(E(x))$ and (ii) a discriminator D .

For (i), we use a U-Net-based architecture, a typical encoder (E) - decoder (G) architecture with skip connections, as a generator and a PatchGAN classifier for the discriminator. Given an input lightness channel $x \in \mathbb{R}^{h \times w \times 1}$, our U-Net-based generator is trained to predict associated a and b color channels (in the CIE *Lab* color space) $y_{ab} \in \mathbb{R}^{h \times w \times 2}$, where h and w are image dimensions. For (ii), we use a convolutional classifier that is trained to classify whether the output image is real (i.e. ground-truth images) or fake (i.e. generated images). Following the work by Mao et al. [28], we adopt the least squares loss function for the discriminator instead of using the sigmoid cross entropy loss function, which may lead to the vanishing gradients problem during the learning process. Our loss functions for the generator and the discriminator are as follows:

$$\begin{aligned} \mathcal{L}_D(x, y_{ab}) &= \frac{1}{2} \mathbb{E}_{x, y_{ab} \sim \mathbb{D}} [(D(x, y_{ab}) - a)^2] \\ &\quad + \frac{1}{2} \mathbb{E}_{x \sim \mathbb{D}} [(D(x, G(E(x))) - b)^2] \\ \mathcal{L}_G(x) &= \mathbb{E}_{x \sim \mathbb{D}} [(D(x, G(E(x))) - c)^2] \end{aligned} \tag{1}$$

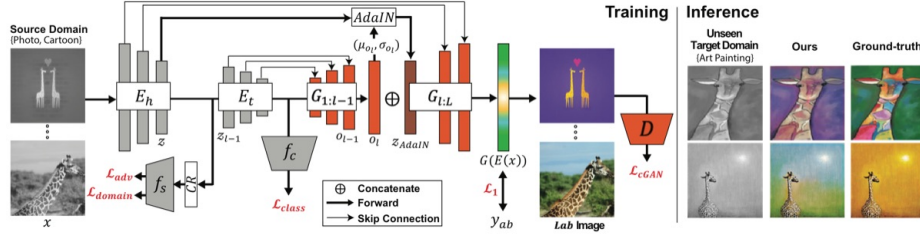


Fig. 2. An overview of our proposed generalizable colorization model. Our model is built upon the existing pix2pix [17] model that utilizes U-Net-based encoder as a generator and a PatchGAN classifier as a discriminator. Our model takes a grayscale image and predicts its color channels. During training, to improve the model’s ability to generalize well to unseen target domains, we propose to use the following three regularizations: (i) Object Class Classifier f_c , (ii) Texture-biased Domain Classifier f_s , and (iii) Content Feature Statistics Transfer via AdaIN (adaptive instance normalization).

where we use the a - b coding scheme for the discriminator and a and b are the labels for real data and fake data, respectively. We use \mathbb{D} to denote the training data distribution. We use c to denote the value that G wants D to believe for fake data. In our implementation, we use $a = 1$, $b = 0$, and $c = 1$. In literature, mixing the GAN objective with a traditional loss function, such as \mathcal{L}_2 distance, is advantageous for being stable during learning process. Thus, we add the following \mathcal{L}_1 distance rather than \mathcal{L}_2 , which often makes the output blurred.

$$\mathcal{L}_1(x, y_{ab}) = \mathbb{E}_{x, y_{ab} \sim \mathbb{D}} [\|y_{ab} - G(E(x))\|_1] \quad (2)$$

Concretely, our loss function for training the conditional GANs is as follows:

$$\mathcal{L}_{cGAN} = \mathcal{L}_D(x, y_{ab}) + \mathcal{L}_G(x) + \lambda_{\text{color}} \mathcal{L}_1(x, y_{ab}) \quad (3)$$

where we use a hyperparameter λ_{color} to control the strength of the traditional loss function.

3.2 Learning Domain-invariant Contents Features

Unlike the human vision system that largely focuses on contents (i.e. shapes) for object recognition, recent studies show that ConvNets have a strong inductive bias towards image texture [2,11,13,30]. Moreover, models trained for the colorization task inherently exhibit a strong bias towards texture, which is an essential prior to a high-quality image generation. Such texture information often varies across different domains than the content, and thus makes models more sensitive to domain shift. We empirically observe this domain shift effect in colorization as shown in Fig. 3, where all conventional colorization models generally fail to generalize to unseen target domains.

From recent work [30], we constrain the head encoder E_h from learning domain-invariant information (i.e., more content-biased and texture-invariant

representation) via an adversarial framework. We use a content randomization (CR) module that interpolates contents feature statistics between different examples (regardless of their object class), i.e. it replaces the content of the input with the randomized content through AdaIN.

Formally, given an input image x_i and a randomly-chosen x' in same batch, we first obtain corresponding latent representations z and z' from the head encoder E_h , respectively. Given the channel-wise means $\mu(z)$ and standard deviations $\sigma(z)$ as style representation, we apply AdaIN to the content of z' with the texture of z .

$$CR(z, z') = \sigma(z) \cdot \left(\frac{z' - \mu(z')}{\sigma(z')} \right) + \mu(z) \quad (4)$$

Such content-randomized representation is then fed into a domain classifier f_s , which needs to correctly predict its source domain given texture-biased representations. Thus, this classifier f_s is trained by minimizing a domain classification loss $\mathcal{L}_{\text{domain}}$:

$$\mathcal{L}_{\text{domain}} = -\mathbb{E}_{x, y_s \sim \mathbb{D}} \left[\sum_{s=1}^S y_s \log f_s(CR(z, z'))_s \right] \quad (5)$$

where S is the number of source domains and $y_i \in \{0, 1\}^S$ is the one-hot domain label.

Our head encoder E_h is then trained by minimizing the following adversarial loss $\mathcal{L}_{\text{adv}} = -\lambda_{\text{adv}} \mathcal{L}_{\text{domain}}$ where λ_{adv} is a hyperparameter to control the strength of \mathcal{L}_{adv} .

Emphasizing Semantic Information We also add an object class predictor f_c that consumes image representations from the tail encoder E_t and performs the object recognition task to regularize the model to learn semantic features useful for our generator G to generate better quality images. Formally, we add the cross entropy loss $\mathcal{L}_{\text{class}}$ given an image x and its class label y_c .

$$\mathcal{L}_{\text{class}}(x, y_c) = -\mathbb{E}_{x, y_c \sim \mathbb{D}} \left[\sum_{c=1}^C y_c \log f_c(E_t(E_h(x)))_c \right] \quad (6)$$

where \mathbb{D} is the training data distribution and C is the number of class categories.

3.3 Transferring Domain-invariant Features

In the previous section, we discussed how we train our encoder to learn content-biased representations by applying a content randomization (CR) module as well as a domain classifier. We observe, however, our generator G exhibits color bias towards source domain during training as the network is forced to learn “source-domain style-sensitive color information”. This is achieved using skip connections between encoder and decoder (see Fig. 2). In particular, statistical

color adjustment is required to generate source-domain style-sensitive color for realistic colorization from the encoder’s content-biased feature. So, Instead of using a simple concatenation via skip connection, we use a style transfer technique using an AdaIN – i.e. given content information from z , we transfer style feature statistics of o_l from the intermediate layer of G . We empirically observe that such a “content” skip connection allows the generator to use such content information directly from the encoder during synthesizing output images. We also observe such a skip connection is critical to improving the colorization performance in the unseen target domains. We summarize our results in Experiment section.

Formally, we first obtain a representation o_l by concatenating the output z from E_h and the latent representation o_l from the l -th layer of G : i.e. $o_l = G_l(o_{l-1} \oplus z_{l-1})$ where G_l is the l -th layer of G . Given the representations z and o_l , we apply AdaIN to the content of z with the style of o_l :

$$z_{\text{AdaIN}} = \sigma(o_l) \cdot \left(\frac{z - \mu(z)}{\sigma(z)} \right) + \mu(o_l) \quad (7)$$

where z_{AdaIN} can be interpreted as a style-transferred representation of z . We concatenate z_{AdaIN} with o_l and feed into the next layer of G : $o_{l+1} = G_{l+1}(o_l \oplus z_{\text{AdaIN}})$. In our experiment, we set $l = 5$. Details of the relevant experiments are provided in the supplementary material.

4 Experiments

4.1 Implementation and Evaluation Details

Implements Details Following [17], we use the same architectural choices for the generator and (color) discriminator. For our G , we use a U-Net-based architecture as a backbone, and for our (color) D , we use a convolutional PatchGAN classifier. The model architectures for the f_s and the f_c are based on the same architecture as that of our E_t . Note that domain labels are finally computed followed by an average pooling layer and three fully connected layers. We train our model using an Adam optimizer [19] for approximately 50 epochs. The batch size is set to 128 and the learning rate to 0.001. Our implementation is based on PyTorch.

Dataset We evaluate the effectiveness of the proposed method on the publicly available PACS [23] and Office-Home [37] benchmark datasets. The PACS dataset contains over 10k images from four diverse domains: Photo, Art Painting, Cartoon, and Sketch. This dataset is particularly useful in domain generalization research as it provides a bigger domain shift than existing photo-only benchmarks. As the Sketch domain does not provide color information, we exclude it from the experiment. This PACS dataset provides seven object categories: dog, elephant, giraffe, guitar, horse, house, and person. We split examples from training domains in the ratio 8:2 (training:validation) and test on the entire held-out domain. Note that we use the best-performed model on validation for testing.

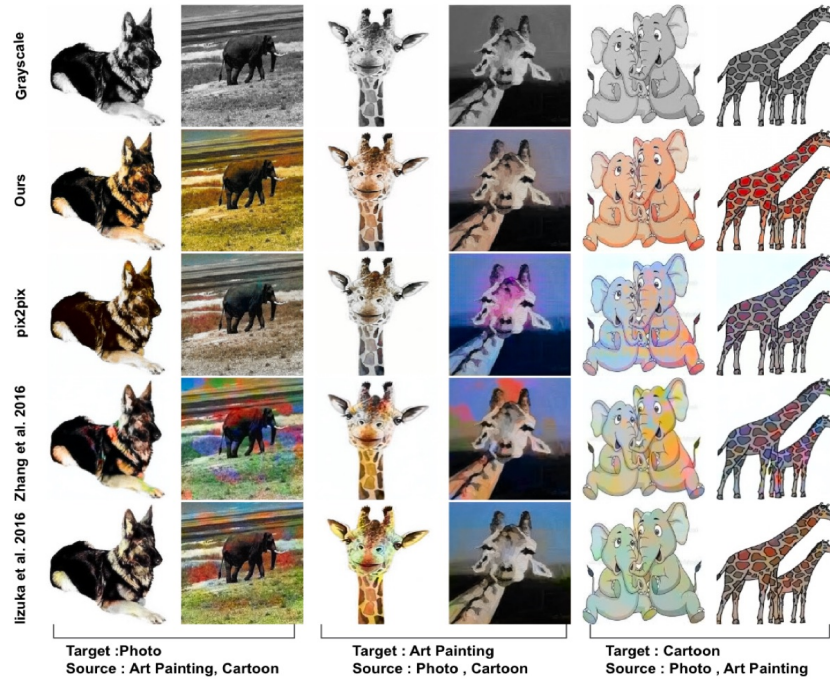


Fig. 3. Colorization performance comparison with conventional colorization approaches. Target and source domains are listed in the bottom row. Data: PACS.

We also use the Office-Home dataset, which contains over 15k images from four domains: Art, Clipart, Product, and Real-World. We exclude the Product domain from our experiment owing to its lack of color information. This dataset provides 65 object categories.

Evaluation Metrics Evaluation of the quality of colorized images is known to be challenging. We first use the following four widely-used quantitative metrics: peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), image quality metrics (IQM), and Frechet Inception distance (FID). The first two metrics compute the pixel-wise distances between the ground-truth and synthesized images—thus quantifying the similarity of colors. Unlike PSNR and SSIM, IQM only uses the a and b color channels of the image in the CIE Lab color space to quantify the image quality in terms of colorfulness, sharpness, and contrast. Thus, IQM is a suitable metric for the colorization task [1]. FID measures the distance between latent image representations for the ground-truth and synthesized images. We use an ImageNet-pretrained Inception v3 model to extract such feature vectors for FID. Note that we also finetuned the Inception v3 model with the corresponding domains to remove the negative effect of domain shift when computing FID.

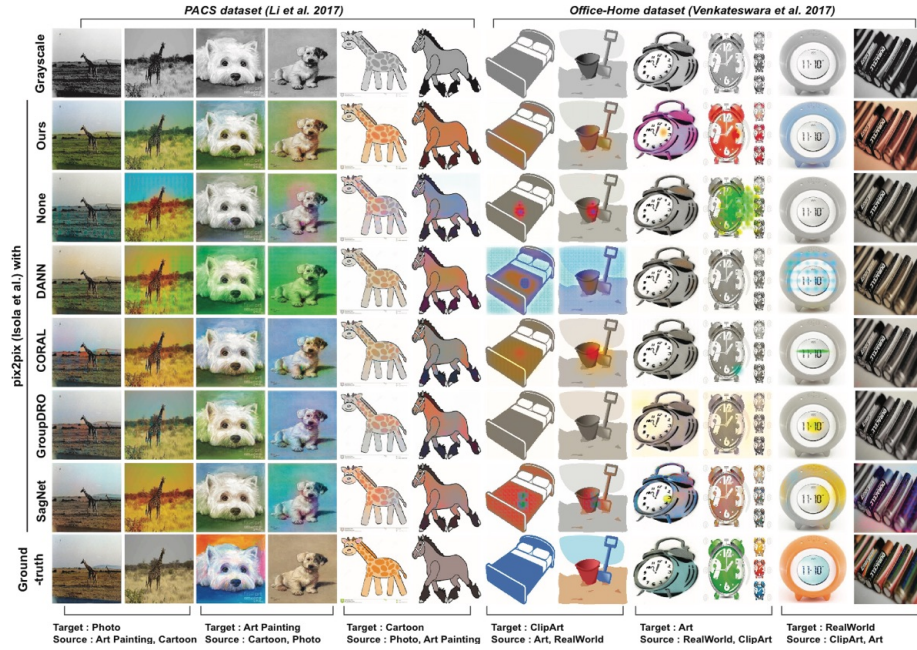


Fig. 4. Qualitative colorization performance comparison with four alternative domain generalization techniques. All models are built upon our baseline pix2pix [17] architecture and we add regularization losses to improve the model’s generalization power. We provide more diverse examples in the supplemental material. We used two datasets: PACS (see 1st-6th columns) and Office-Home (see 7th-12th columns).

However, these metrics often fail to capture visual realism. Although the aforementioned four quantitative metrics are frequently used in colorization tasks, further discussions of clear quantitative performance indicators of colorization are still ongoing [1,3,20,44,45]. To address the shortcomings of any of the above individual evaluations, we further evaluate our model using a user study.

4.2 Effect of Domain Shift in Colorization

We first investigate the effect of domain shift in the colorization task with the following three landmark colorization models: Zhang et al. [43], Iizuka et al. [15], and pix2pix [17]. We use the leave-one-out setting, i.e. a pre-selected single domain is used as a test domain and the others as training domains. We use the PACS and Office-Home datasets for this experiment. As shown in Fig. 3, we observe that all models generally fail to generate successful colorized images; in fact, they often show a failure to capture long-range color consistency, a sepia-tone on complex scenes, and confusion between red and blue information. This is further confirmed by our quantitative analysis in terms of the four image quality metrics:

Table 1. Colorization performance comparison in terms of four image quality evaluation metrics. An average value across domains is reported. To observe any performance degradation in the domain generalization setting, we also compare each model with the non-domain generalization (non-DG) setting, i.e. models are trained using the same target domain. Data: PACS [23] and Office-Home [37].

Models	PACS [23]				Office-Home [37]			
	PSNR \uparrow	SSIM \uparrow	IQM \uparrow	FID \downarrow	PSNR \uparrow	SSIM \uparrow	IQM \uparrow	FID \downarrow
A. Zhang et al. [43]	30.56	0.65	1.89	30.78	32.25	0.74	1.63	17.77
B. A w/ non-DG setting	33.48	0.79	1.84	12.96	-	-	-	-
C. Iizuka et al. [15]	30.81	0.80	1.91	23.52	32.67	0.79	1.48	17.25
D. C w/ non-DG setting	33.49	0.82	1.84	14.18	-	-	-	-
E. pix2pix [17]	30.38	0.63	1.80	25.28	32.69	0.74	1.63	19.41
F. E w/ non-DG setting	31.56	0.71	1.89	21.28	-	-	-	-
G. E + DANN [10]	30.12	0.59	1.68	25.39	31.07	0.67	1.53	22.53
H. E + CORAL [35]	30.59	0.65	1.84	24.63	32.15	0.73	1.57	16.09
I. E + GroupDRO [33]	30.51	0.65	1.82	25.93	31.98	0.75	1.47	18.35
J. E + SagNet [30]	30.67	0.68	1.81	27.46	31.92	0.67	1.61	18.69
K. E + Ours	30.71	0.66	1.88	24.92	32.29	0.73	1.52	14.92

PSNR, SSIM, IQM, and FID. In Table 1, a large degradation with these metrics is observed for all models. This clearly indicates that the current colorization models do not generalize well to novel target domains and additional treatment is needed to deal with such a domain shift. We provide more detailed numerical values and failed examples for each domain in the supplementary material.

4.3 Effect of Domain Generalization Techniques

To improve the generalization power of the colorization models, we explore the effect of applying some existing domain generalization techniques. Note that these models originally focused on the object recognition task, not on the colorization task. Thus, for a fair comparison, all models are based on the pix2pix model that generally shows better colorization performance than alternatives. We then implement the core idea of each domain generalization technique from DANN [10], CORAL [35], GroupDRO [33], and SagNet [30]. Note that other techniques may also be applicable, but we leave them as future work. As summarized in Table 1 (in the 7th-10th rows for alternatives and in the 11th row for ours), all domain generalization techniques (except DANN [10]) generally provide better results than our baseline pix2pix model (5th row) in terms of four image quality evaluation metrics. In particular, ours generally outperforms others and shows better

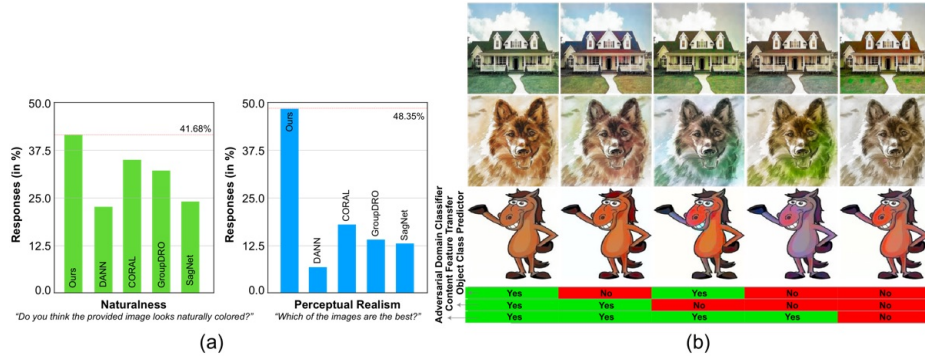


Fig. 5. (a) Evaluation of perceptual realism by a user study. Participants were asked to answer two questions for evaluating naturalness (left) and perceptual realism (right). (b) Ablation study results between variants of our models with and without the following three modules: Adversarial Domain Classifier, Content Feature Transfer, and Object Class Predictor.

capability in generalizing to unseen target domains. This indicates that considering both domain-invariant contents and source domain-specific color information is suitable for the generalized colorization task.

In Fig. 4, we provide some random examples of the generated colorized images. Models are trained on the PACS (1st-6th columns) and Office-Home (7th-12th columns) datasets. For a better comparison, we provide the corresponding ground-truth images in the 8th row as well as the results from our baseline model (i.e. pix2pix) in the 3rd row. Not surprisingly, the baseline model generates lower-quality images; they often generate gray-tone images on the Office-Home dataset and fail to capture long-range color consistency. Applying existing domain generation techniques (as observed by comparing the 3rd vs. 4-7th rows) generally provides better-quality colorized images than our baseline, but their generated images still show limitations as they show confusion between blue and red channels. This failure is more apparent in the Office-Home dataset, which is more challenging for colorizing unseen-domain images.

Ours, however, generally shows comparable or better performance against alternative domain generalization techniques. Our quantitative analysis, as presented in Table 1, confirms that our proposed method outperforms others in terms of four image quality evaluation metrics. Our qualitative analysis in Fig. 4 further confirms that exploiting domain-invariant contents and source domain-specific color features enables the proper colorization of images (see 2nd row vs. others). In Office-Home benchmark, neither baseline nor existing domain generalization techniques succeeded in colorizing the object in novel domains, whereas our model could realistically color these objects.

Table 2. Ablation study results between variants of our models. An average value is reported. Data: PACS.

Models	λ_{adv}	PSNR \uparrow	SSIM \uparrow	IQM \uparrow	FID \downarrow	Acc.
A. Ours	1.0	30.71	0.66	1.88	24.92	38.5
B. Ours ($\lambda_{adv} = 0.5$)	0.5	30.61	0.66	1.84	25.48	36.8
C. Ours ($\lambda_{adv} = 0.1$)	0.1	30.63	0.65	1.93	27.06	36.4
D. A - Object Class Predictor f_c	1.0	30.67	0.67	1.84	25.23	-
E. A - Content Feature Transfer	1.0	29.77	0.62	1.76	28.25	-
F. A - Content Feature Transfer - Object Class Predictor f_c	1.0	30.42	0.63	1.82	26.10	-
G. F - Adversarial Domain Classifier f_s (baseline)	-	30.38	0.63	1.80	25.28	-
H. Ours w/ Style Feature Transfer	1.0	30.52	0.67	1.77	30.10	-

4.4 Evaluation of Perceptual Realism by User Study

Evaluating colorized images is generally difficult using a set of automatic metrics, such as PSNR. As our main goal is to make the colorized images that are more compelling to human observers, we set up a user study, in which we show participants synthesized colors for an image, and ask them to answer the following two questions: (i) Naturalness: do you think the provided image looks naturally colored? (ii) Perceptual Realism: which of the images are the best? Images were randomly sampled from each domain on the PACS and Office-Home datasets. For this user study, 34 participants were recruited and each participant answers overall 360 questions and submitted 12,240 votes. A detailed explanation is provided in the supplementary material.

As shown in Fig. 5 (a), ours significantly outperforms alternatives (including existing domain generalization approaches) with a large gap in both questions. 41.68% (2,551 out of 6,120 votes) of colorized images by ours were perceived as naturally colored. For evaluating perceptual realism, 49.26% (2,959 out of 6,120 votes) of colorized images by ours was chosen as the best-colored images among five approaches: DANN, CORAL, GroupDRO, SagNet, and ours. This number is significantly higher than all compared approaches, and these results validate the effectiveness of the proposed method.

4.5 Ablation Study

Effect of Object Class Predictor Recall that our model uses an object class predictor that takes domain-invariant content features as an input and performs object recognition tasks to encourage the model to learn semantic features necessary to predict their classes. We observe in Fig. 5 (b) that this predictor is particularly useful to generate more saturated images (compare A vs. E columns). As shown in Table 2, our quantitative analysis also confirms that performance is generally degraded without the use of object class predictor as a regularization (compare Model A vs. D).

Effect of Transferring Domain Feature Statistics Recall that we transfer content feature statistics from encoder to decoder as a skip connection followed

by AdaIN, which is widely used in the style transfer task. To see its effect, we conduct an ablation study and we observe in Table 2 that colorization performance is generally degraded as we turn the content feature transfer off (see Model A vs. E), which is probably because our generator is easily biased towards source domain-specific color information. This is more apparent in our qualitative analysis as shown in Fig. 5 (b) where the color for source domain is better recovered (see 1st vs. 3rd columns).

Further, to verify our motivation behind transferring content feature statistics, we evaluate a variant of ours where we apply AdaIN to the content of o_l with the style of z (instead of using the content of z with the style of o_l). As expected, constraining the generator towards using content-biased features generally degrades the overall performance in colorization (see model A vs. H in Table 2). This may confirm that the generator needs to be source domain color-biased network, while the encoder needs to be content-biased network. We provide examples in the supplementary material.

Effect of λ_{adv} We observe in Table 2 that decreasing λ_{adv} generally degrades the colorization evaluation scores (see models A-C), which may confirm that learning content-biased representations is beneficial for the model to generalize well to unseen domains. This trend is more apparent in object class classification performance (see the rightmost column in the table). Ours outperforms other alternative domain generalization approaches (Acc. of SagNet: 36.5).

5 Conclusion

We proposed an innovative fully automatic colorization algorithm that can generalize well to unseen target domains. Built upon a conditional GAN-based colorization deep neural network architecture, we proposed three modules to learn domain-invariant content-biased encoder and source domain-specific color generator. A skip connection between them transfers rich information about content feature statistics. Our extensive experiments demonstrate ours generally outperforms alternative domain generalization techniques, and our user study further confirmed this. To the best of our knowledge, we are the first to explore the effect of domain shift in the colorization task.

Acknowledgement

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2021-0-00994, Sustainable and robust autonomous driving AI education / development integrated platform). J. Kim was supported by the MSIT (Ministry of Science and ICT), Korea, under the ICT Creative Consilience program (IITP-2022-2020-0-01819) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation)

References

1. Anwar, S., Tahir, M., Li, C., Mian, A., Khan, F.S., Muzaffar, A.W.: Image colorization: A survey and dataset. arXiv preprint arXiv:2008.10774 (2020)
2. Baker, N., Lu, H., Erlikhman, G., Kellman, P.J.: Deep convolutional networks do not classify based on global object shape. PLoS computational biology **14**(12), e1006613 (2018)
3. Cao, Y., Zhou, Z., Zhang, W., Yu, Y.: Unsupervised diverse colorization via generative adversarial networks. In: Joint European conference on machine learning and knowledge discovery in databases. pp. 151–166. Springer (2017)
4. Carlucci, F.M., D’Innocente, A., Bucci, S., Caputo, B., Tommasi, T.: Domain generalization by solving jigsaw puzzles. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2229–2238 (2019)
5. Charpiat, G., Hofmann, M., Schölkopf, B.: Automatic image colorization via multimodal predictions. In: European conference on computer vision. pp. 126–139. Springer (2008)
6. Cheng, Z., Yang, Q., Sheng, B.: Deep colorization. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 415–423 (2015)
7. Chia, A.Y.S., Zhuo, S., Gupta, R.K., Tai, Y.W., Cho, S.Y., Tan, P., Lin, S.: Semantic colorization with internet images. ACM Transactions on Graphics (TOG) **30**(6), 1–8 (2011)
8. D’Innocente, A., Caputo, B.: Domain generalization with domain-specific aggregation modules. In: German Conference on Pattern Recognition. pp. 187–198. Springer (2018)
9. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of the International Conference on Machine Learning (ICML). pp. 1126–1135. PMLR (2017)
10. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V.: Domain-adversarial training of neural networks. The journal of machine learning research **17**(1), 2096–2030 (2016)
11. Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F.A., Brendel, W.: Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. arXiv preprint arXiv:1811.12231 (2018)
12. Gupta, R.K., Chia, A.Y.S., Rajan, D., Ng, E.S., Zhiyong, H.: Image colorization using similar images. In: Proceedings of the 20th ACM international conference on Multimedia. pp. 369–378 (2012)
13. Hermann, K., Chen, T., Kornblith, S.: The origins and prevalence of texture bias in convolutional neural networks. Advances in Neural Information Processing Systems **33**, 19000–19015 (2020)
14. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1501–1510 (2017)
15. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Let there be color! joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. ACM Transactions on Graphics (ToG) **35**(4), 1–11 (2016)
16. Ironi, R., Cohen-Or, D., Lischinski, D.: Colorization by example. Rendering techniques **29**, 201–210 (2005)
17. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017)

18. Kim, M., Byun, H.: Learning texture invariant representation for domain adaptation of semantic segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 12975–12984 (2020)
19. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
20. Lei, C., Chen, Q.: Fully automatic video colorization with self-regularization and diversity. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 3753–3761 (2019)
21. Levin, A., Lischinski, D., Weiss, Y.: Colorization using optimization. In: *ACM SIGGRAPH 2004 Papers*, pp. 689–694 (2004)
22. Li, C., Wand, M.: Precomputed real-time texture synthesis with markovian generative adversarial networks. In: *European conference on computer vision*. pp. 702–716. Springer (2016)
23. Li, D., Yang, Y., Song, Y.Z., Hospedales, T.M.: Deeper, broader and artier domain generalization. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (2017)
24. Li, D., Yang, Y., Song, Y.Z., Hospedales, T.M.: Learning to generalize: Meta-learning for domain generalization. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 32 (2018)
25. Li, H., Pan, S.J., Wang, S., Kot, A.C.: Domain generalization with adversarial feature learning. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 5400–5409 (2018)
26. Li, Y., Tian, X., Gong, M., Liu, Y., Liu, T., Zhang, K., Tao, D.: Deep domain generalization via conditional invariant adversarial networks. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. pp. 624–639 (2018)
27. Liu, X., Wan, L., Qu, Y., Wong, T.T., Lin, S., Leung, C.S., Heng, P.A.: Intrinsic colorization. In: *ACM SIGGRAPH Asia 2008 papers*, pp. 1–9 (2008)
28. Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Paul Smolley, S.: Least squares generative adversarial networks. In: *Proceedings of the IEEE international conference on computer vision*. pp. 2794–2802 (2017)
29. Muandet, K., Balduzzi, D., Schölkopf, B.: Domain generalization via invariant feature representation. In: *Proceedings of the International Conference on Machine Learning (ICML)*. pp. 10–18. PMLR (2013)
30. Nam, H., Lee, H., Park, J., Yoon, W., Yoo, D.: Reducing domain gap by reducing style bias. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 8690–8699 (2021)
31. Nazeri, K., Ng, E., Ebrahimi, M.: Image colorization using generative adversarial networks. In: *International conference on articulated motion and deformable objects*. pp. 85–94. Springer (2018)
32. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. pp. 234–241. Springer (2015)
33. Sagawa, S., Koh, P.W., Hashimoto, T.B., Liang, P.: Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. *arXiv preprint arXiv:1911.08731* (2019)
34. Su, J.W., Chu, H.K., Huang, J.B.: Instance-aware image colorization. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 7968–7977 (2020)
35. Sun, B., Saenko, K.: Deep coral: Correlation alignment for deep domain adaptation. In: *European conference on computer vision*. pp. 443–450. Springer (2016)

36. Vapnik, V.: Statistical learning theory new york. NY: Wiley (1998)
37. Venkateswara, H., Eusebio, J., Chakraborty, S., Panchanathan, S.: Deep hashing network for unsupervised domain adaptation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5018–5027 (2017)
38. Vitoria, P., Raad, L., Ballester, C.: Chromagan: Adversarial picture colorization with semantic class distribution. In: The IEEE Winter Conference on Applications of Computer Vision. pp. 2445–2454 (2020)
39. Wang, Y., Li, H., Kot, A.C.: Heterogeneous domain generalization via domain mixup. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 3622–3626. IEEE (2020)
40. Welsh, T., Ashikhmin, M., Mueller, K.: Transferring color to greyscale images. In: Proceedings of the 29th annual conference on Computer graphics and interactive techniques. pp. 277–280 (2002)
41. Xu, M., Zhang, J., Ni, B., Li, T., Wang, C., Tian, Q., Zhang, W.: Adversarial domain adaptation with domain mixup. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 6502–6509 (2020)
42. Yan, S., Song, H., Li, N., Zou, L., Ren, L.: Improve unsupervised domain adaptation with mixup training. arXiv preprint arXiv:2001.00677 (2020)
43. Zhang, R., Isola, P., Efros, A.A.: Colorful image colorization. In: European conference on computer vision. pp. 649–666. Springer (2016)
44. Zhang, R., Zhu, J.Y., Isola, P., Geng, X., Lin, A.S., Yu, T., Efros, A.A.: Real-time user-guided image colorization with learned deep priors. arXiv preprint arXiv:1705.02999 (2017)
45. Zhao, J., Han, J., Shao, L., Snoek, C.G.: Pixelated semantic colorization. International Journal of Computer Vision pp. 1–17 (2019)