Spatial Feature Combination for Generative Creativity: A Case Study of Bionic Design

Simiao Yu[1], Hao Dong[1], Pan Wang[1], Chao Wu[2], Yike Guo[1]
[1]Imperial College London [2]Zhejiang University
[1]{simiao.yu13,hao.dong11,pan.wang15,y.guo}@imperial.ac.uk
[2]{chao.wu}@zju.edu.cn

Abstract

We aim to achieve generative creativity by learning to combine spatial features of images from different domains. We focus on shape-oriented bionic design as an ideal case study: a target object (e.g. a floor lamp) is designed to contain shape features of biological source objects (e.g. flowers), resulting in creative biologically-inspired design. We propose DesignGAN, a conditional GAN based architecture with several designated losses for this task. We demonstrate that our proposed method can successfully generate creative images of bionic design.

1 Introduction

In computer vision, achieving generative creativity (i.e. the generation of new and creative images via feature combination) is a long-term goal. For example, in image style transfer [3, 9, 4], creative images can be generated by composing the features of existing content images and style images in a novel manner. In this work, we aim to achieve generative creativity by learning to combine spatial features of images from different domains. Bionic design [8, 17], in which a biologically-inspired object is created by combining the features of a target design object with those of biological source objects, offers an ideal context for our study. We focus on *shape-oriented* bionic design: given an input image of the design target, our objective is to generate biologically-inspired images that 1) maintain the shape features of the input image, 2) contain the shape features of images from the biological source domain, 3) remain plausible and diverse. Essentially, shape-oriented bionic design exhibits generative creativity, because it can be seen as composing the spatial features of design target images and biological source images into creative images that never existed before.

The challenges are threefold. First, the task is of unsupervised learning, since the nature of creative design implies that there is no or very few available images of biologically-inspired design. In our case, we only have unpaired data of design target images and biological source images. Second, there should be multiple ways of integrating features of biological source images into the given design target image. In other words, bionic design is a one-to-many generation process, and the learned generative model should be able to achieve this variation. Third, the generated biologically-inspired design should preserve key features of input design target image and biological source images, which requires the model to be able to select and combine the salient features of different sources.

2 Model

Following the assumption and problem formulation (detailed in Appendix A), we propose DesignGAN (Figure 1), a conditional generative adversarial networks (cGAN) [5, 14] based framework for bionic design, with various enhancements designed to resolve the aforementioned challenges. First, the generator takes as input both an image and a latent variable sampled from a prior distribution, which

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

enables the generation of diverse images. This is implemented by the introduction of an encoder and a latent loss. Second, our approach employs both a cycle loss [19, 11, 18] and a regression loss to help maintain the salient features of the design target. Last, an adversarial loss is used to integrate the salient features of biological source images into the input design target image. More details can be found in Appendix B.

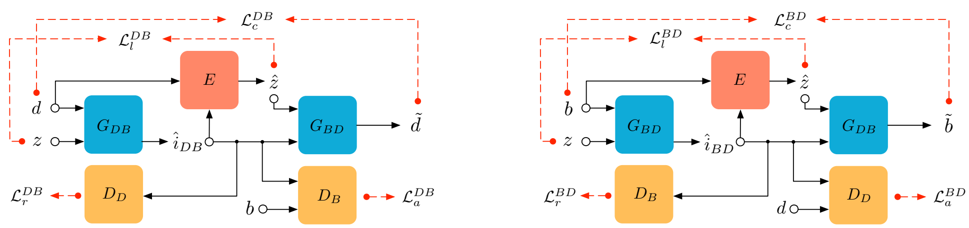

Figure 1: Schema of our proposed DesignGAN model. It is composed of two generators (G_{DB} and G_{BD}), two discriminators (D_B and D_D) and one encoder (E). Loss components include latent loss ($\mathcal{L}_l^{DB}$ and $\mathcal{L}_l^{BD}$), cycle loss ($\mathcal{L}_c^{DB}$ and $\mathcal{L}_c^{BD}$), regression loss ($\mathcal{L}_r^{DB}$ and $\mathcal{L}_r^{BD}$) and adversarial loss ($\mathcal{L}_a^{DB}$ and $\mathcal{L}_a^{BD}$). At inference stage, G_{DB} is used to generate varied biologically-inspired images $\hat{i}_{DB}$, given an input design target image d and an input noise vector z.

3 Results

Figure 2 illustrates the generated shape-oriented bionic design by our proposed DesignGAN (more experimental details can be found in Appendix C). We maintained the same value of the latent variable for the corresponding three generated images for each group of generation. DesignGAN is capable of generating creative and diverse biologically-inspired images that contain the combined spatial representations of both input design target image and biological source images.

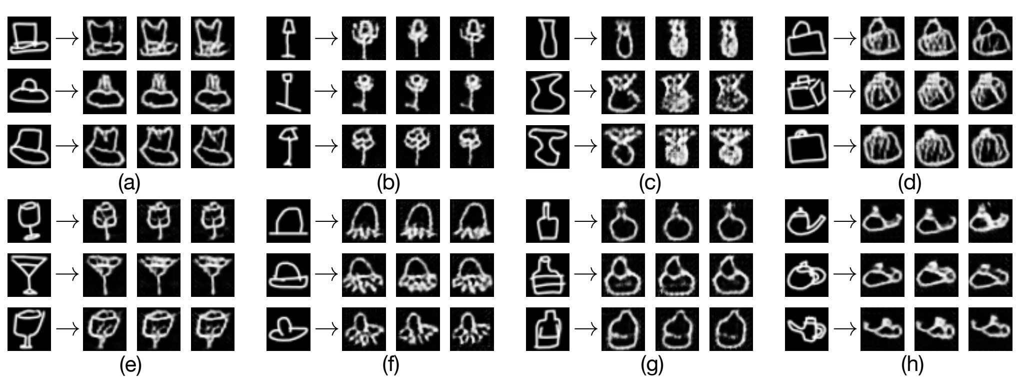

Figure 2: Qualitative results of our proposed DesignGAN for shape-oriented bionic design. (a) Hat + rabbit. (b) Floor Lamp + flower. (c) Vase + pineapple. (d) Suitcase + onion. (e) Wine glass + flower. (f) Hat + octopus. (g) Wine bottle + pear. (h) Teapot + whale.

Figure 3 shows the generated biologically-inspired design images by linearly interpolating the input latent variable z. The smooth semantic transitions verify that our model learns useful representations for the bionic design problem, rather than a simple memorisation of the training samples [15].

Figure 3: Results of interpolating input latent variable. (a) Suitcase + onion. (b) Floor lamp + flower.

References

[1] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: a system for large-scale machine learning. In *OSDI*, 2016.

[2] H. Dong, A. Supratak, L. Mai, F. Liu, A. Oehmichen, S. Yu, and Y. Guo. TensorLayer: a versatile library for efficient deep learning development. In *ACM-MM*, 2017.

[3] A. A. Efros and W. T. Freeman. Image quilting for texture synthesis and transfer. In *SIGGRAPH*, 2001.

[4] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *CVPR*, 2016.

[5] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *NIPS*, 2014.

[6] D. Ha and D. Eck. A neural representation of sketch drawings. In *ICLR*, 2018.

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

[8] M. Helms, S. S. Vattam, and A. K. Goel. Biologically inspired design: process and products. *Design Studies*, 30(5):606 – 622, 2009.

[9] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin. Image analogies. In *SIGGRAPH*, 2001.

[10] S. Ioffe and C. Szegedy. Batch normalization: accelerating deep network training by reducing internal covariate shift. In *ICML*, 2015.

[11] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim. Learning to discover cross-domain relations with generative adversarial networks. In *ICML*, 2017.

[12] D. Kingma and J. Ba. Adam: a method for stochastic optimization. In *ICLR*, 2014.

[13] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. P. Smolley. Least squares generative adversarial networks. In *ICCV*, 2017.

[14] M. Mirza and S. Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[15] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative Adversarial Networks. In *ICLR*, 2016.

[16] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb. Learning from simulated and unsupervised images through adversarial training. In *CVPR*, 2017.

[17] L. Shu, K. Ueda, I. Chiu, and H. Cheong. Biologically inspired design. *CIRP Annals*, 60(2):673 – 693, 2011.

[18] Z. Yi, H. Zhang, P. Tan, and M. Gong. DualGAN: unsupervised dual learning for image-to-image translation. In *ICCV*, 2017.

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

Appendix

A Assumption and Problem Formulation

The problem of bionic design can be formulated as follows. Given a design target domain D containing samples $\{d_k\}_{k=1}^{M} \in D$ (e.g. floor lamps) and a biological source domain B containing samples $\{b_k\}_{k=1}^{N} \in B$ (e.g. flowers), we have the corresponding latent spaces of D and B (respectively Z_d and Z_b) that contain the representations of each domain. We denote the data distribution of D and B as $p(d)$ and $p(b)$. We then make two key assumptions of the bionic design problem: 1) there exists an "intermediate" domain I containing the generated objects of biologically-inspired design $\{\hat{i_k}\}_{k=1}^{O} \in I$, and 2) the corresponding latent space of I (denoted as Z) contains the *merged* representations of those from Z_d and Z_b, as illustrated in Figure 4.

Based on these two assumptions, the objective of bionic design is to learn a generating function $G_{DB} : D \times Z \to I$, such that the generative distribution matches the distribution of I (denoted as $p(i)$). Since in our case we do not have any existing samples from I, it is impossible to explicitly learn

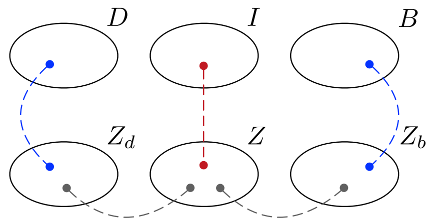

Figure 4: Our assumption of the bionic design problem.

such generative distribution. Nonetheless, we could still learn it in an implicit fashion via real data distributions $p(d)$ and $p(b)$, and the careful design of the model architecture. This is where generative creativity comes from. Also note that G_{DB} takes as input the latent variable $z \in Z$ sampled from the distribution $p(z)$, the requirement of variations for bionic design is satisfied directly: multiple samples based on a single d can then be generated by sampling different z from $p(z)$.

B Methods

Each loss component of DesignGAN is detailed as follows.

Adversarial loss. We employ two sources of adversarial loss $\mathcal{L}_a^{DB}(G_{DB}, D_B)$ and $\mathcal{L}_a^{BD}(G_{BD}, D_D)$ that respectively enforce the outputs of G_{DB} and G_{BD} to match the empirical data distribution $p(b)$ and $p(d)$, as an approach to integrate corresponding features to the generated images.

$$\begin{aligned}
\mathcal{L}_a(G_{DB}, G_{BD}, D_B, D_D) = \mathcal{L}_a^{DB}(G_{DB}, D_B) + \mathcal{L}_a^{BD}(G_{BD}, D_D),& \\
\mathcal{L}_a^{DB}(G_{DB}, D_B) = \mathbb{E}_{b\sim p(b)}[\log D_B(b)] + \mathbb{E}_{d\sim p(d), z\sim p(z)}[\log(1 - D_B(G_{DB}(d, z)))],& \\
\mathcal{L}_a^{BD}(G_{BD}, D_D) = \mathbb{E}_{d\sim p(d)}[\log D_D(d)] + \mathbb{E}_{b\sim p(b), z\sim p(z)}[\log(1 - D_D(G_{BD}(b, z)))],&
\end{aligned} \tag{1}$$

where D_B and D_D are discriminators that distinguish between generated and real images from B and D.

Cycle loss. The problem of bionic design requires the generated images to maintain the features of the input design target. In other words, the generated image should still be recognised as in the class of the design target. For the shape-oriented bionic design problem, it simply implies that the generated images should resemble the input images to a large extent. After all, it would be unreasonable to generate biologically-inspired images that in turn share no relationship to the input design target image. We apply cycle loss $\mathcal{L}_c^{DB}$ and $\mathcal{L}_c^{BD}$ to constrict the generators G_{DB} and G_{BD} to retain the shape representations of the input images:

$$\begin{aligned}
\mathcal{L}_c(G_{DB}, G_{BD}) &= \mathcal{L}_c^{DB}(G_{DB}, G_{BD}) + \mathcal{L}_c^{BD}(G_{BD}, G_{DB}), \\
\mathcal{L}_c^{DB}(G_{DB}, G_{BD}) &= \mathbb{E}_{d\sim p(d), z\sim p(z)}[\|G_{BD}(G_{DB}(d, z), E(G_{DB}(d, z), d)) - d\|_2^2], \\
\mathcal{L}_c^{BD}(G_{BD}, G_{DB}) &= \mathbb{E}_{b\sim p(b), z\sim p(z)}[\|G_{DB}(G_{BD}(b, z), E(b, G_{BD}(b, z))) - b\|_2^2],
\end{aligned} \tag{2}$$

where we employ L2 norm in the loss. The inclusion of cycle loss makes our model optimised in a dual-learning fashion [19, 11, 18]: we introduce an auxiliary generator G_{BD} and train all the generators and discriminators jointly. After training, only G_{DB} will be used for bionic design purpose.

Regression loss. The cycle loss enforces the generated images to maintain the shape features of the input image only. Another way of maintaining the design target features is to simultaneously force the generated images to contain key features of the design target domain, which directly makes the generated images recognised as the class of the design target. We therefore introduce the regression loss L_r^{DB} and L_r^{BD} imposed by the discriminator D_D and D_B. L_r^{DB} and L_r^{BD} respectively constricts G_{DB} and G_{BD} to maintain representations from the domain of input images. Note that in such a situation D_D and D_B are employed as a regression function only, without competing with the generators as the adversarial loss does. This is why in Figure 1 there is only one input to D_D and D_B when referring to $\mathcal{L}_r$.

$$
\begin{aligned}
\mathcal{L}_r(G_{DB}, G_{BD}) &= \mathcal{L}_r^{DB}(G_{DB}) + \mathcal{L}_r^{BD}(G_{BD}), \\
\mathcal{L}_r^{DB}(G_{DB}) &= \mathbb{E}_{d\sim p(d), z\sim p(z)}[\log(1 - D_D(G_{DB}(d, z)))], \\
\mathcal{L}_r^{BD}(G_{BD}) &= \mathbb{E}_{b\sim p(b), z\sim p(z)}[\log(1 - D_B(G_{BD}(b, z)))].
\end{aligned}
\tag{3}
$$

Latent loss. We employ a unified encoder E and a latent loss to model the variation of the bionic design problem:

$$
\begin{aligned}
\mathcal{L}_l(G_{DB}, G_{BD}, E) &= \mathcal{L}_l^{DB}(G_{DB}, E) + \mathcal{L}_l^{BD}(G_{BD}, E), \\
\mathcal{L}_l^{DB}(G_{DB}, E) &= \mathbb{E}_{d\sim p(d), z\sim p(z)}[\|E(G_{DB}(d, z), d) - z\|_1], \\
\mathcal{L}_l^{BD}(G_{BD}, E) &= \mathbb{E}_{b\sim p(b), z\sim p(z)}[\|E(b, G_{BD}(b, z)) - z\|_1].
\end{aligned}
\tag{4}
$$

The encoder E of DesignGAN encodes a pair of images from each domain (either $(\hat{i}_{DB}, d)$ or $(b, \hat{i}_{BD})$) into the latent space Z of domain I, which corresponds to our assumption of the bionic design problem (Appendix A). The latent loss is computed by the L1 norm distance between the generated latent variable $\hat{z}$ and the input noise vector z, which forces the model to generate diverse output images.

Full objective. The full objective function of our model is:

$$
\begin{aligned}
\min_{\{G_{DB}, G_{BD}, E\}} \max_{\{D_B, D_D\}} \mathcal{L}(G_{DB}, G_{BD}, E, D_B, D_D) &= \lambda_a \mathcal{L}_a(G_{DB}, G_{BD}, D_B, D_D) \\
&+ \lambda_c \mathcal{L}_c(G_{DB}, G_{BD}) + \lambda_r \mathcal{L}_r(G_{DB}, G_{BD}) + \lambda_l \mathcal{L}_l(G_{DB}, G_{BD}, E),
\end{aligned}
\tag{5}
$$

where we employ λ_a, λ_c, λ_r and λ_l to control the strength of individual loss components.

C Experimental Details

Dataset. We evaluated our models on "Quick, Draw!" dataset [6] that contains millions of simple grayscale drawings of size 28×28 across 345 common objects. It is an ideal dataset for the shape-oriented bionic design problem due to its contained images of great variation and complexity. We selected eight pairs of domains of design targets and biological sources as the varied bionic design problems, including hat + rabbit, floor Lamp + flower, vase + pineapple, suitcase + onion, wine glass + flower, hat + octopus, wine bottle + pear, teapot + whale. We randomly chose 4000 images from each domain of the domain pairs for training.

Network architecture. For the generator networks, we adopted the encoder-decoder architecture. The encoder contained three convolutional layers and the decoder had two transposed convolutional layers. Six residual units [7] were applied after the encoder. The latent vector was spatially replicated and concatenated to the input image. The discriminator networks contained four convolutional layers. For the encoder network, the two input images were concatenated and encoded by three convolutions and six residual units. We employed ReLU activation in the generators and encoder, and leaky-ReLU activation in the discriminators. Batch normalisation [10] was implemented in all networks.

Training details. The networks were trained for 120 epochs using Adam optimiser [12] with a learning rate of 0.0001 and a batch size of 64. The learning rate was decayed to zero linearly over the last half number of epochs. Due to the distinct complexity of images from different domains, the values of λ_a, λ_c, λ_r and λ_l and dimension of latent variable z were set independently for each of the domain pairs. We used the objective functions of Least Squares GAN [13] to stabilise the learning process. The discriminator was updated using a history of generated images, as proposed in [16], in order to alleviate the model oscillation problem [19]. We appled random horizontal flipping and random ±15 degree rotation to the training images, which were further resized to 32×32 before being fed into the models. The implementation was in TensorFlow [1] and TensorLayer [2].