Palette2Interior Architecture: From Syntactic and Semantic Colour Palettes to Generative Interiors with Deep Learning

Immanuel Koh[1]

[1] Singapore University of Technology and Design, Singapore, Singapore
Immanuel_koh@sutd.edu.sg

Abstract. Colour palettes have long played a significant role in not only capturing design ambience (e.g., as mood boards), but more significantly, in translating an abstract intuition into an explicit ordering mechanism for design representation and synthesis, whether it is in the discipline of graphic design, interior design or architectural design. Might this difficult process of design synthesis from a low-dimensional colour input domain to a high-dimensional spatial design output domain be computationally mapped? Using today's generative adversarial networks (GANs), the paper aims to investigate this plausibility, and in doing so, hoping to envision an AI-augmented design workflow and tooling. Newly-created datasets are made procedurally and used to train three different types of deep learning models in the specific context of generating living room interior layouts. The results suggest that a combination of syntactic and semantic generative processes is necessary for a critical appropriation of such AI models

Keywords: Machine Learning, Artificial Intelligence, Deep Neural Networks, Colour Palette, Interior Design

1 Introduction

Although there exist many online colour palette generators for designers to create coherent colour schemes given any image or photograph inputs, such as those based on clustering (Lertrusdachakul et al., 2019), those used for image editing (Grogan et al., 2018) and those used for image re-colourisation (Cho et al., 2017), the inverse is not the case. That is, given any colour palette as input, might one generate a design as its output? This one-directional computational extraction of key colours from inspirational images are most commonly used by graphic/web/interior designers to quickly build a colour palette that could then guide the subsequent stages of their design process, such as spatial compositions. It is thus reasonable to think that an inverse version of the same tool would be of little use. However, one might recall artists like Paul Klee and Sol LeWitt who had used limited set of discrete colour bands as abstraction of semantics and abstraction of syntax respectively. For example, in the striated composition of Klee's 1929 *Fire in the Evening* where the red opaque rectangle semantically suggests a flaming fire in the desert during the evening, and in LeWitt's 1992 *Bands of Color in Four Directions (Within A Square)* where the painting is the literal manifestation of its title and nothing more -- both are inherently generative, though visually reductive as perceptual inputs. Yet, such a process is akin to the inverse version of those aforementioned colour palette extractors. The research aims to explore the question of whether such inversion might be appropriated for the architectural design generative process. The paper will proceed to first lay out the processes used in the creation of the datasets and the implementation of the deep learning models proposed in 'Methodology' section, followed by an illustration and evaluation of the experimental outputs in the 'Results' section, before ending with a discussion on the current and future work in the 'Discussion' section.

2 Methodology

Using interior design as a domain of investigation, the proposed AI-augmented design workflow uses low-dimensional colour inputs in yielding high-dimensional spatial outputs. In generating high-dimensional photorealistically composed interior views, the proposed 'syntactic' palette2interior AI model based on the original GAN (Goodfellow et al., 2014) and the convolutional variant of GAN (Radford et al., 2015), and more specifically cycleGAN (Zhu et al., 2017) or cycle-consistent adversarial network architecture, is trained from scratch with simple inputs of 5 rectilinear colour bands in varying orders and orientations; while the proposed 'semantic' palette2interior AI model also based on the pix2pix or cGAN (Isola et al., 2018) or conditional generative adversarial network architecture trained with complex inputs of 150 possible colours in organic shapes in varying sizes. As part of the generative exploration, a third AI model based on the StyleGAN (Karras et al., 2019) architecture is trained to obtain a latent space for creating a new synthetic dataset of the original interior views. The study consists of dataset preparation, model architecture selection, model training and evaluation, and lastly, design and deployment of an interface application for model inference. The first dataset contains web-scraped interior photographs of living rooms, the second dataset contains automated annotations of the first dataset using a pretrained semantic segmentation model called PSPNet or pyramid scene parsing network (Zhao et al., 2017), and the third dataset contains permutated bands of colours extracted from the first dataset using Google's online colour palette generator.

2.1 Datasets

A total of three corresponding datasets have been created for the experiments, with 80% used as training set and 20% as test set. The first dataset forms the basis on which to derive the other remaining two auto-annotated datasets. Preliminary research in identifying potential sources of imagery containing living room interiors are conducted. Two websites have been found particularly suitable in providing the necessary quality (e.g., without watermarks) and minimal quantity that can be webscraped for the training of the proposed deep learning models. By searching the phrase 'living room interior' on *Unsplash* and the phrase 'interior design living rooms' on *Google Images*, thousands of images are scraped (Fig. 1).

The second dataset is derived by using a colour palette extraction procedure to obtain the five main colours of any given image from the first dataset (Fig. 1). The initial approach is to use the KMeans clustering algorithm to extract the top five colours present in any given image based on their count values. However, the colour palette extracted in this manner does not in fact reflect the actual perceptual colour palette with which a designer or artist would have used in reality. Therefore, the online colour palette extraction capability of *Google Arts & Culture* website is eventually used which indeed yields superior results in effectively representing the five most perceptually dominant colours. The colour information is first stored as hexadecimal value (e.g., #FFFFFF as white) before converting into RGB values (e.g., (255,255,255) as white). The five colours are further converted into an image represented by five discrete horizontal colour bands. Since the order of the horizontal colour bands has no intrinsic meaning in relation to the extraction process, five (instead of all) permutations with/without rotations are further generated as a data augmentation procedure to form an expanded dataset that can be used for training the deep learning models.

The third dataset is derived by using a semantic image segmentation model to obtain the semantic labels to every pixel in any given image from the first dataset. The DeepLabV3 model (Chen et al., 2017) is initially tested on the first dataset, but performed poorly as it fails to recognize several indoor features, such as tables and shelves. The reason is that it has been pre-trained with the Cityscapes dataset (Cordts et al., 2016) which contains only outdoor scenes and streetscapes, rather than indoor environments. Therefore, the pyramid scene parsing network or PSPNet model (Zhao et al., 2017) is eventually used which yields superior results due to the fact that it is pre-trained on the ADE20K dataset (Zhou et al., 2018) that contains both outdoor and indoor scenes.

The last step of the data preparation and annotation generation process is to explore and visualize all three datasets to gain an understanding of their respective distributions and correlations. Fig. 2 shows one of the visualizations that clusters all the indoor interior images from the first dataset based on their respective colour palatte from the second dataset. This is done by first converting each palette's 5 RGB values into 15-dimensional features, projecting them onto a 2-dimensional plot using principal component analysis (PCA), and finally performing a KMeans clustering with a k-value of 3. It can be observed that cluster 0 (shown as purple points in the scatter plot) gathers interiors with a darker colour palette (i.e., 'Warm' tones), cluster 1 (shown as green points in the scatter plot) gathers interiors with a lighter and warmer colour palette (i.e., 'Neutral' tones), and cluster 3 (shown as yellow points in the scatter plot) gathers interiors with a lighter and cooler colour palette (i.e., 'Cool' tones). In other words, the extracted colour palettes generated for the second dataset is indeed reflective of the first dataset with which they are derived from.

The actual datasets to be used for training the proposed deep learning models consists of different pairing combinations among the three datasets, that is, the first dataset (living room interior imagery) combined with the generated interior images from the second and third datasets, a pair of the first dataset with the second dataset (colour palette), and a pair of the first dataset with the third dataset (semantically segmented interior). The next section will describe the proposed deep learning models used to train on these different datasets.

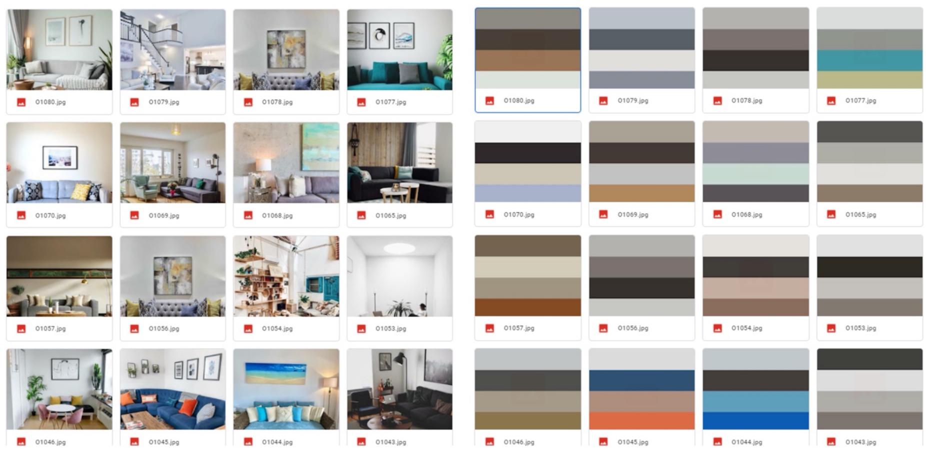

Figure 1. (LEFT) Dataset 1 containing images of living room interiors; (RIGHT) Dataset 2 generated from Dataset 1 and contains the corresponding colour palettes with the top 5 colours as horizontal bands. Source: Author, 2022.

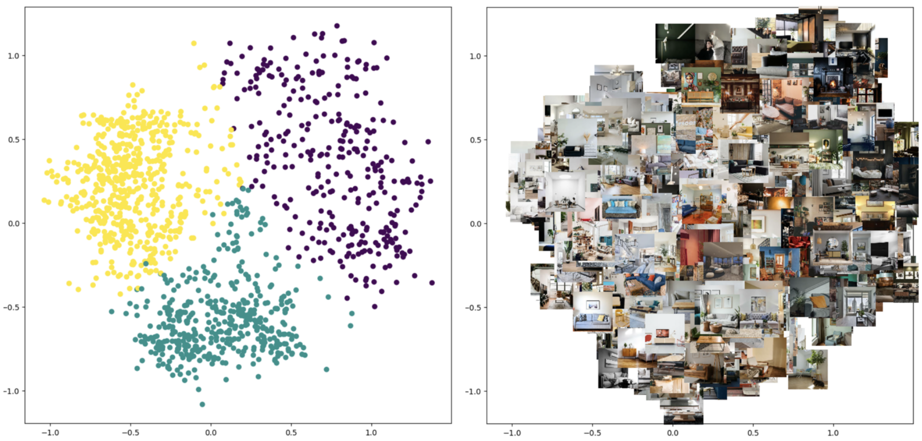

Figure 2. (LEFT) KMeans clustering based on the features of colour palettes from Dataset 2; (RIGHT) Plot of the corresponding living room interior imagery. Source: Author, 2022**.**

2.2 Models

A total of three different deep learning models have been implemented and trained with the aforementioned datasets. All of which are based on variants of the architecture of the generative adversarial networks (GANs). The first model 'syntactic' palette2interior model is an unpaired image-to-image translation convolutional neural network model called cycleGAN (Zhu et al., 2017) that is trained by very loosely pairing the living room interior imagery with its colour palette imagery. The task is to generate a new photorealistic living room interior composition output from a 5-colour palette imagery input (with or without reordering and reorientating the colour bands with permutations and rotation respectively). The model is trained for a total of 195 epochs and restored at epoch 174 without data augmentation (Fig. 3). That is, using only the original extracted horizontal bands of colours. Another model is trained for 123 epochs and restored at epoch 105 with data augmentation (Fig. 4). That is, using reordering of the horizontal bands and reorientation of the bands via orthogonal rotations as the expanded dataset.

The second model 'semantic' palette2interior model is a paired image-to-image translation convolutional neural network model called pix2pix (Isola et al., 2018) that is trained by strictly pairing the living room interior imagery with its semantically segmented imagery. The task is to generate a new photorealistic living room interior composition output from any organically drawn shapes whose pixel colours belong to the set of pixel labels used for training. The model is trained for a total of 210 epochs (Fig. 5).

Finally, as part of the generative exploration, a third AI model based on the StyleGAN (Karras et al., 2019) architecture is trained to obtain a latent space for creating a new synthetic dataset of the original interior views. Transfer-learning from a StyleGAN model pre-trained on bedrooms is used to avoid having to train from scratch. The model is trained for a total of 1000 epochs and reaches a final FID (Frechet Inception Distance) Score of 25.38.

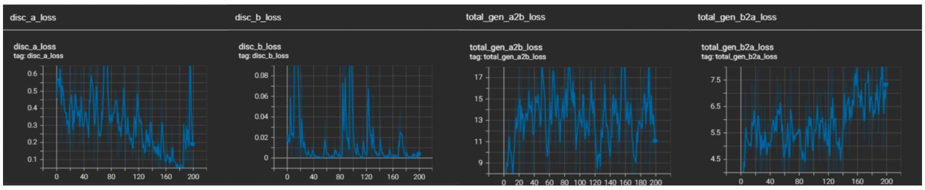

Figure 3. Plots showing the losses (discriminator losses in column 1 & 2 and generator losses in column 3 & 4) during the training of the 'syntactic' palette2interior model with the pairing of the living room interior imagery with its colour palette imagery as dataset. Source: Author, 2022.

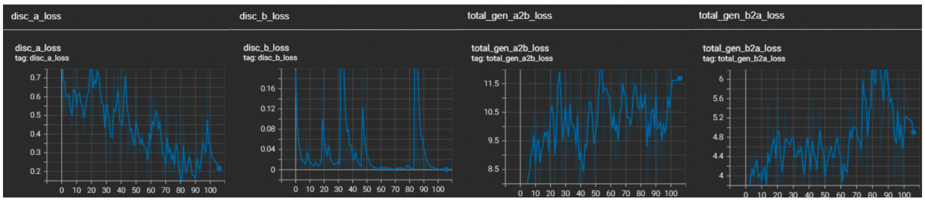

Figure 4. Plots showing the losses (discriminator losses in column 1 & 2 and generator losses in column 3 & 4) during the training of the 'syntactic' palette2interior model with the pairing of the living room interior imagery with its colour palette imagery (with augmentation via reordering and reorientation) as dataset. Source: Author, 2022.

Figure 5. Plots showing the losses (discriminator loss in column 1 and generator losses in column 2, 3 & 4) during the training of the 'semantic' palette2interior model with the pairing of the living room interior imagery with its semantically segmented imagery as dataset. Source: Author, 2022.

3 Results

The results from all three models will now be evaluated. In Fig. 6, despite using different colour palettes as inputs to the 'syntactic' palette2interior model, the generated spatial layouts do not seem to vary significantly.

However, it is apparent that the colour palette will still influence the overall colour layout within the interior imagery. More significantly, the reorientated version of the same colour palette will alter both the spatial and colour layouts of the generated outputs. In Fig. 7, even with different colour palettes as inputs to the 'syntactic' palette2interior model, it is observed that the generated spatial layouts do not vary significantly, the overall colour layout do vary substantially. Here, the reorientated colour palette can likewise significantly alter the outputs. In Fig. 8, since the colours are assigned spatially as conditioning pixel label inputs to the 'semantic' palette2interior model, the generated outputs show relatively accurate, realistic and interpretable interior design layouts. In fact, the generated resolution of larger objects improves with further training. However, there is no mechanism to control the desired generated colour palette layout other than the spatial layout. The former is only possible with the previous 'syntactic' palette2interior model. Fig. 9 shows the generated outputs of the third model trained with the StyleGAN architecture. Since the pre-trained model used for the transfer learning is based on a dataset of bedrooms, it can be observed that there are moments when beds morph into sofas in the generated latent walk video.

In general, based on the results from the syntactic model, it is observed that even an abstract spatial composition as that of 5 colour bands could contribute directly to the generated interior spatial composition. However, the placement order of the 5 colour bands within the same palette will only generate differing interior spatial composition if they are perceived as significantly different-looking palettes as images. Since the semantic model is more compositionally descriptive, it is relatively more successful in generating spatially plausible outputs. An interface (Fig. 10) is built to better visualise how such a 'palette2interior' application might be used in real-time with 2 different interactive modes, namely, assigning 5 colours to form an orthogonally composed input palette for the syntactic model, and colouring on a canvas to form an organically composed input palette for the semantic model.

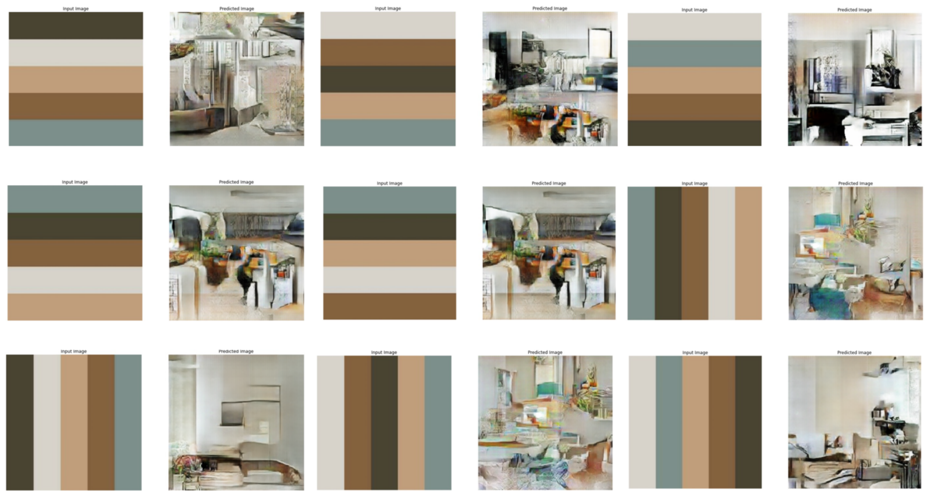

Figure 6. Although the same colour palette when reordered might yield similar spatial layouts with variation mainly in colour layout, it is observed that the position of the darkest colour band can dramatically influence the overall generated outputs. The reorientated colour palette can likewise significantly alter the outputs. Source: Author, 2022**.**

Figure 7. With different colour palettes as inputs, it is observed that the generated spatial layouts do not vary significantly, although the colour palette will still influence the overall colour layout. The reorientated colour palette can likewise significantly alter the outputs. Source: Author, 2022**.**

Figure 8. Given that the colours now spatially represent the object semantics, the generated outputs show relatively accurate, realistic and interpretable interior design configurations. Source: Author, 2022.

Figure 9. Samples of the generated outputs of the third model. Transfer-learning is used to quicken the training with the original dataset of living room interiors. See its latent walk video at https://vimeo.com/729465453/3de4c42727. Source: Author, 2022

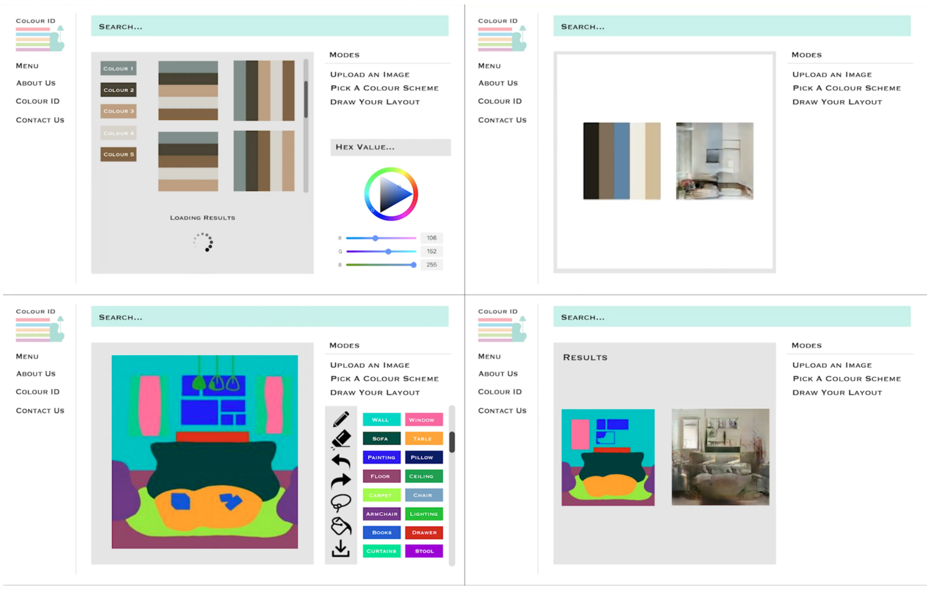

Figure 10. The interface of the proposed 'palette2interior' application with 2 different interactive modes, namely, assigning 5 colours to form an orthogonally composed input palette for the syntactic model, and colouring on a canvas to form an organically composed input palette for the semantic model. See the interactive video recording at https://vimeo.com/698194423/7152d7b5db Source: Author, 2022.

4 Discussion

The paper has contributed to an alternative understanding of the ways in which deep neural networks might be appropriated in inverting the traditional conception of colour palettes as a means towards a semantically and syntactically higher form of design synthesis. There are several limitations to the current work, such as the model inability to accommodate non-orthogonal colour palette representations as shown in Fig. 11. Future work will also combine the 'syntactic' model with the 'semantic' model to better control the colour and spatial layouts simultaneously in a single AI model.

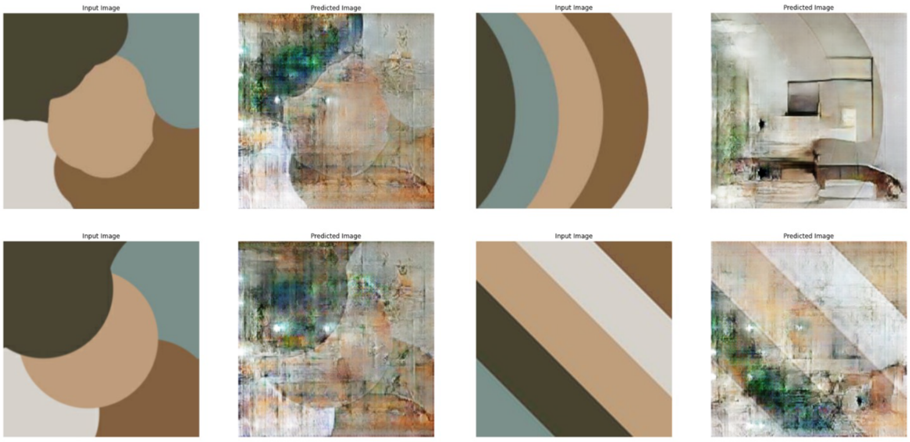

Figure 11. The limitation of the current 'syntactic' palette2interior model in accommodating non-orthogonal colour palette representations can be observed here. Source: Author, 2022**.**

References

Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (2017). *Rethinking Atrous Convolution for Semantic Image Segmentation* (arXiv:1706.05587; Version 3). arXiv. https://doi.org/10.48550/arXiv.1706.05587

Cho, J., Yun, S., Lee, K., & Choi, J. Y. (2017). PaletteNet: Image Recolorization with Given Color Palette. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1058–1066. https://doi.org/10.1109/CVPRW.2017.143

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., & Schiele, B. (2016). *The Cityscapes Dataset for Semantic Urban Scene Understanding* (arXiv:1604.01685). arXiv. https://doi.org/10.48550/arXiv.1604.01685

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Networks. *ArXiv:1406.2661 [Cs, Stat]*. http://arxiv.org/abs/1406.2661

Grogan, M., Hudon, M., McCormack, D., & Smolic, A. (2018). Automatic Palette Extraction for Image Editing. *Undefined*. https://www.semanticscholar.org/paper/Automatic-Palette-Extraction-for-Image-Editing-Grogan-Hudon/fb1b4d4c49c28f2ba3134c07dd190ad891572357

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2018). Image-to-Image Translation with Conditional Adversarial Networks. *ArXiv:1611.07004 [Cs]*. http://arxiv.org/abs/1611.07004

Karras, T., Laine, S., & Aila, T. (2019). A Style-Based Generator Architecture for Generative Adversarial Networks. *ArXiv:1812.04948 [Cs, Stat]*. http://arxiv.org/abs/1812.04948

Lertrusdachakul, T., Ruxpaitoon, K., & Thiptarajan, K. (2019). Color Palette Extraction by Using Modified K-means Clustering. *2019 7th International Electrical Engineering Congress (IEECON)*, 1–4. https://doi.org/10.1109/iEECON45304.2019.8938867

Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *ArXiv:1511.06434 [Cs]*. http://arxiv.org/abs/1511.06434

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid Scene Parsing Network. *ArXiv:1612.01105 [Cs]*. http://arxiv.org/abs/1612.01105

Zhou, B., Zhao, H., Puig, X., Xiao, T., Fidler, S., Barriuso, A., & Torralba, A. (2018). *Semantic Understanding of Scenes through the ADE20K Dataset* (arXiv:1608.05442). arXiv. https://doi.org/10.48550/arXiv.1608.05442

Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *ArXiv:1703.10593 [Cs]*. http://arxiv.org/abs/1703.10593