香港城市大學
City University of Hong Kong

Department of Computer Science

BSCCS Final Year Project Report 2021-2022

21CS073

Deep learning model for analysing visual styles of buildings

(Volume __ of __)

Student Name : **DAI, Chenyue**

Student No. : **55670822**

Programme Code : **BSCEGU4**

Supervisor : **Dr LAU, Rynson W H**

Date : **January 31, 2022**

For Official Use Only

Student Final Year Project Declaration

I have read the project guidelines and I understand the meaning of academic dishonesty, in particular plagiarism and collusion. I hereby declare that the work I submitted for my final year project, entitled:

Deep learning model for analysing visual styles of buildings

does not involve academic dishonesty. I give permission for my final year project work to be electronically scanned and if found to involve academic dishonesty, I am aware of the consequences as stated in the Project Guidelines.

Student Name: DAI, Chenyue Signature:

Student ID: 55670822 Date: March 30, 2022

Abstract

The visual styles, categorized primarily by their elements, including patterns, shapes, and textures, are constantly an essential topic in research, education, and the creation of art and architecture. Styles form from design and creation, while they, in turn, influence succeeding design, and creation. Thus, the study of architectural style has its unique significance in reflecting and exploring new ideas from the history of architecture and even society, as well as inspiring creative, experimental practices.

The rapid development of computer vision and image processing algorithms and models has seen success in all walks of life in recent years. However, the topic of architectural styles in computer vision and pattern recognition is still relatively constrained in the domain of algorithm development with academic output. Thus, this project aims at developing an interactive application with advanced deep learning methods. It seeks to provide classification and style transfer functions for a broader range of audiences for education, research, and artistic creation.

The system built in this project consists of three parts, including a classification module, a style transfer module, and a frontend interactive webpage. The classification module is trained on ResNet50, designed by He et al. A large portion of this project focuses on the style transfer module, which is achieved by a combination of depth map generation and pix2pix image translation with transfer learning the training process. The interactive webpage is set up by HTML, javascript, and flask.

Acknowledgements

I would like to express my sincere appreciation to all the people who supported me to keep working on this challenging task till the end.

Firstly, I would like to thank my supervisor, Dr. Rynson Lau, who actively discussed with me and gave me suggestions for starting this project and gave me valuable guidance and criticism on reports.

Secondly, I would like to thank the professors and coordinators from the Department of Computer Science at the City University of Hong Kong. They have given me this great opportunity to turn my unique idea into practice, which has become a starting point of my new career as an interdisciplinary art and architecture student. Also, I would like to appreciate the whole department's effort to keep everything on track, including all the resources, during such a difficult pandemic. Besides, the past two years of study in the Department of Computer Science prior to this project have definitely provided fundamental support to me.

Thirdly, I would like to thank my close friends Mr. Chen Zirui, Mrs. Wu Shanshan, and Mr. Li Jiaqian for motivating me throughout the whole year of progress and providing some simple assistance like manual filtering the dataset. I am also grateful to my architecture advisor Mr. Zhang Yunyu for encouraging me to think creatively.

Last but not least, I would like to thank City University of Hong Kong for giving me such an excellent platform for studying, exchanging thoughts, and exploring my own ideas. Without these valuable four years here, I may not be able to step onto my dream path to MIT Architecture, towards a career as an architect and the top of the world. Also, I would like to appreciate my beloved family. They are always my strongest backing during my study here, especially in various tough times.

Table of Contents

1. Introduction

1.1 Background

The visual style is constantly an essential topic in the historical research and education of art [1]. As a form of art, architectural works can also be categorized primarily by their elements, including patterns, shapes, and textures. Considering the cultures and civilizations that architecture is often closely connected to, Such studies and practices have their unique significance in reflecting on the history of architecture and even society [2].

In recent years, the rapid development of computer vision and image processing algorithms and models has brought this technology to all walks of life and achieved significant success in countless problems involving images and visual elements [14]. Thus, we can foresee huge potential when such advanced technology engages with the long-discussed topic of architectural styles.

1.2 Motivation

The rapid development of computer vision and image processing algorithms and models has seen success in all walks of life in recent years. However, the topic of architectural styles in computer vision and pattern recognition is still relatively constrained in the domain of algorithm development with academic output.

The classification works on architectures, which involve analysis of visual features, have been a topic in the computer vision community for years [4…7], while deep learning models for general computer vision and image processing tasks have also seen a rapid improvement. This project investigates the performance of newer networks on this task.

Meanwhile, style transfer is also a popular topic that attracts researchers interested in computer vision and artists pursuing advanced tools for creating works. Thus, this project poses a new question: can style transfer be applied to architectural styles? Also, since the realistic elements and features of buildings are different from brush strokes and texture information of paintings that are considered as the "style" to be transferred in most of the previous works [11], inspirations are expected for both improvements of style transfer models and design of architectures.

Overall, this project aims at developing an interactive application with advanced deep learning methods. It seeks to provide classification and style transfer functions for a broader range of audiences for education, research, and artistic creation.

1.3 Objectives

The project is expected to serve as a creative tool for architectural learning and artistic creation with classification and synthesis functions.

First, the project investigates the architectural styles worldwide and selects common or representative ones for building the dataset. Second, current models for classification are tested and improved for the widened range of styles. Its accuracy is expected to be as high as possible since it will affect the style transfer part. Lastly, the model is expected to transfer styles chosen from the previous model to image recognized successfully for the style transfer part. The result needs to be further analysed and evaluated as acceptable by humans since style transfer brings artistic and distorted output in images.

An interactive website will be set up to realize the following pipeline: 1) Users input photos of buildings. 2) The classification model recognizes the styles of buildings. 3) Users choose new styles existing in the model, and the application transfer the new styles to the photos.

2. Literature Review

2.1 Architectural Styles Dataset

Architectural styles are defined by the characteristics of the buildings. They may be featured by various visual elements, including form, material, construction methods, and decorative components. There are a number of mainstream styles defined by scholars and the general public, primarily based on chronology and regions. However, architectural styles may evolve, spread, and influence each other [1, 2, 3]. Therefore, the classification of architectures is a complicated problem because of the links and branches existing in the styles. On the other side, one architecture may also carry characteristics from multiple styles. For example, the White House carries a combination of Neoclassical and Palladian styles.

In Xu *et al*. 's work [4], they collected an architectural style dataset from Wikimedia's hierarchy of keyword "Architecture_by_style". Twenty-five styles with most images are preserved. However, their choice of the dataset has its weakness and is not ideally suitable for this project. For example, most Achaemenid architectures are highly damaged ruins and piles of stones, which are of little value to extract style. American craftsman style consists of one-to-two-story American domestic houses, while most of the other architectures are larger-scale buildings with high reputation, such as cathedrals, museums, and commercial buildings. Also, most of the Colonial architectures carry a large portion of styles from their mother country.

2.2 Classification of Architectural Styles

The topic of classification of architectural styles has been discussed for a long time. Early work on this topic was proposed by Goel *et al*. [4] to explore the possibility of using SIFT feature extraction, SVM method, and modified BoG method on low-level features. Xu *et al*. [5] later proposed a combination of the Deformable Part-based Model and Multinomial Latent Logistic Regression to analyze higher-level architectural features. Zhao *et al*. [6] improved their work by the Improved Ensemble Projection method to extract the shared characteristics of the same style and differences among different styles, which succeeded in higher accuracy and a more extensive range of styles than Xu *et al*.'s method. Another piece by Yoshimura *et al*. [7] used the DCNN model NASNet to learn and cluster features, which achieves high accuracy but investigates individual architects' works instead of general styles.

Meanwhile, deep learning neural networks for general image classification have been further developed after the published works mentioned above. The increasing number of layers are able to capture features in various levels, including different scales. Among such researches, Simonyan *et al.* [22] proposed a very deep convolutional network, from which VGG16 and

VGG19, two configurations of 16 and 19 layers, became trendy models for a wide range of tasks, including style transfer that will be discussed below. Later, He *et al.* [23] successfully implemented neural networks with much more layers. They tackled vanishing and exploding gradient problems commonly seen in deep networks with skip connections. The model also provides various suggested configurations, including 18 layers, 34 layers, and 50 layers.

2.3 Style Transfer

A series of manually made images from artists in [7] perfectly showed what architectural style transfer is referring to. Unlike painting style transfer, the detailed building components like windows may be inconsistent following the feature of the style, while the overall shapes will be preserved (see Fig. 1).

Figure 1: Bauhaus style (right) being transferred onto Buckingham Palace (left) manually by digital artist. The overall shape preserved while detailed components including windows and roof modified according to Bauhaus's features.

2.3.1 Neural Style Transfer

Neural style transfer is a popular topic, though no significant works are specified for visual architectural styles. One early paper from Gatys *et al.* [9] proposed a neural algorithm to separate and recombine the style and content of images. Later, Gatys *et al.* used the VGG network to extract the object information as content and texture information as style [10] and proposed extended ways to control factors in style transfer, including spatial location, color, and scale [11].

A newer model, adaptive instance normalization (AdaIN), presented by Huang *et al.*, achieved real-time arbitrary style transfer by implementing instance normalization [12]. Chandran *et al.* improved it to compensate for the loss of local geometric structures in the style image by adding a conditioning 2D style filter and a separable pointwise convolution tensor [13].

Li *et al.* proposed a whitening and coloring transform method (WCT) to eliminate the learning process for individual styles and presented a multi-level stylization pipeline [14]. Yoo *et al.* improved their work with the addition of wavelet corrected transfer to achieve more photorealistic results [15].

The third model Avatar-Net by Sheng *et al.,* simultaneously achieved high quality, zero-shot, and real-time with a style decorator for semantic style feature distribution and an hourglass network for multi-scale style adaptation [16].

An *et al.* addressed the content leak phenomenon existing in all models above and presented a new framework ArtFlow, which significantly improved the original models [17].

2.3.2 Conditional Generative Adversarial Networks

Generative models, especially conditional generative adversarial networks (cGAN), are an excellent alternative to neural style transfer in terms of their additional focus on preserving the underlying structure of images.
One early work from Wang *et al.* [18] emphasized the importance of structure and style in image generation and their lack in previous GAN models. In their experiments, they proposed S2-GAN, a GAN model that successfully generated relatively realistic indoor scenes based on surface normal maps.

The following work by Isola *et al.* [19] proposed Pix2pix, which focuses on image-to-image translation and supports a more comprehensive range of tasks. The advantage of their method lies in the automatic learning of mapping and loss functions that enables the application of a generic approach to problems that traditionally require manual analysis of particular functions. The generator of Pix2pix is a U-Net, which enables the model to perform pixel-level image to image translation. The discriminator of the model, PatchGAN, provides adversarial loss in a patched way.

Later, Zhu *et al.* [21] designed Cycle-Consistent Adversarial Network (CycleGAN) for unpaired image-to-image translation. GAN architecture similar to Pix2pix is used in their design, but two of them with reversed translation directions are connected and form a cycle. The design makes use of additional cycle consistency loss to learn translation without paired training images. Liu *et al.* [20] have made a significant attempt in architectural style transfer using an improved CycleGAN. They focused on turning ruins or old buildings into modern buildings and have successfully proved the possibility of applying cGAN model to this kind of problem.

As this project can be converted into an image-to-image translation problem but focuses less on detailed components, a preprocessor to provide input that can lead reasonable mapping, such as semantic segmentation addressed by Tyleˇcek *et al.* [26], is still needed to transfer styles aligning with the content of the input successfully. After investigation, depth map is currently the best reference image, and Yin *et al.*'s work [27] showed a good performance on depth prediction.

Meanwhile, GAN algorithms have seen many successful applications in the field of art and architecture. Pix2pix has been open to the community for creative applications since the early day of its existence [19]. Machine Hallucination by Anadol (see Fig. 2) [24] is one of GAN's most successful and relevant artwork, which performs a style-transfer-like synthesis of American urban scenes. Periaktoi: An A.I. interpretation of the Scaenae Frons by Lefebvre [25] is another example where image synthesis is used for architectural design.

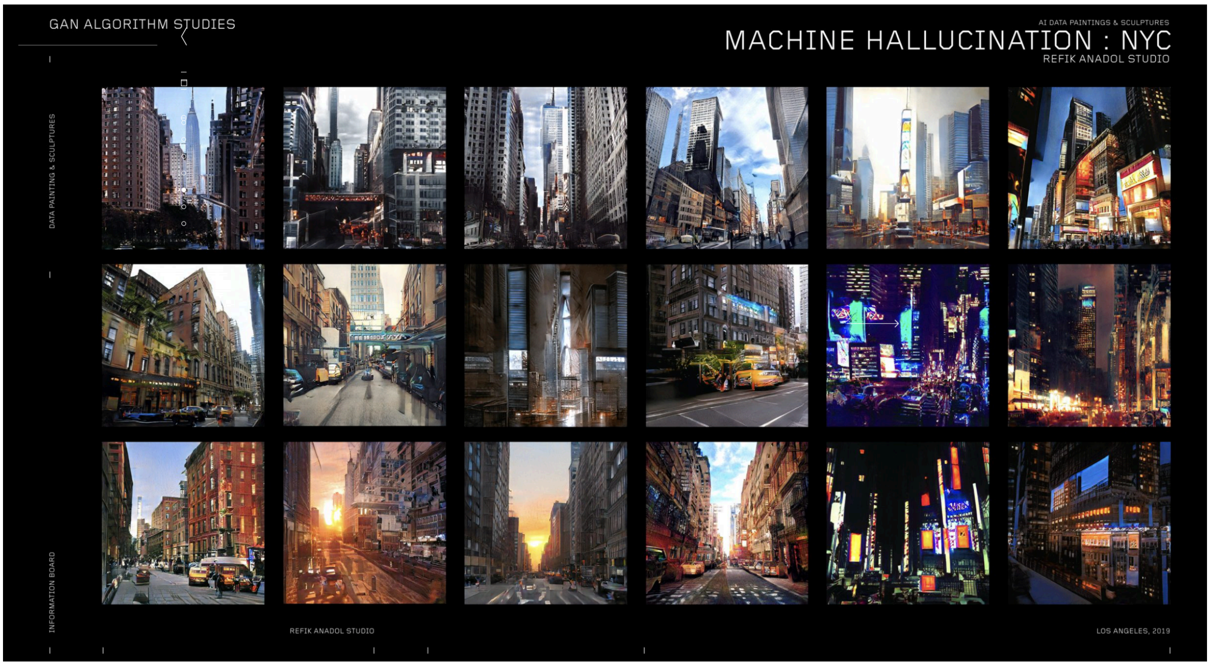

Figure 2: A series of synthesized urban scenes in artwork Machine Hallucination

3. Proposed Design

The system is expected to have a backend that contains class recognition and style transfer models and a front end that handles interaction.

3.1 Classifier Component

The class recognition component is a pre-trained residual neural network that can recognize the architectural style of newly input images from the style set.

More specifically, the model used here is ResNet50, a 50-layer version of the residual neural network designed by He *et al.* The network utilizes skip connections to avoid the problem of vanishing gradient or accuracy saturation problems led by simply increasing the depth of plain network.

3.2 Style Transfer Component

As mentioned above, there are two options for the style transfer component.
The first is a neural style transfer model adapting to architectural style features, mainly including structural features. The model tested is an Artflow enhanced WCT model. The central part of the model consists of a VGG19 encoder, reversed decoder, and whitening and coloring transform part, while Artflow is an unbiased reversible network to avoid content leaks.

The second is a Pix2pix model, which requires an additional preprocessor that processes the input image into the condition map for the model to generate a new image. The chosen preprocessor module is a depth prediction model that generates a depth map from the input image. To achieve image synthesis with different styles, this project utilizes transfer learning. First, the model is trained on the whole dataset of 24720 images so that the network can learn low-level features appearing on all architectures. Then, 16 copies of the trained model are further trained on each style's architecture images. In the end, there are 16 versions of parameters for the generator to synthesize all 16 different styles.

3.3 Dataset

A new dataset is designed and created in this project by scrapping images from Bing Images. In response to the shortcoming of the dataset from Xu *et al.*[4], the involving styles have been reconsidered. Multiple styles are abandoned because of their weak uniqueness or little spread. Two pairs, Bauhaus and International style, Neoclassical and Beaux-art style, are combined since they share a lot of common features. Also, Chinese architecture is added to increase the variety of the dataset.

Since Wikimedia has a limited number and variety of photos, this dataset is collected purely from a scrapping script. In this project, representatives of each style are manually listed first, and the script scraps images by the name of each architecture (see Table 1). Directly scrapping images by style names may lead to a large amount of manual filtering work, the hardest of which may be to determine whether to keep or remove some architectures with mixed or unclear styles. Also, such searching methods may result in only the most widely used photos of some representatives. By contrast, searching an individual architecture may get photos from both official archives and tourists, and it is also much easier to identify a particular building than style from irrelevant images.

The images collected are manually filtered to exclude irrelevant images, watermarked photos, interior photos, night scenes, or components of the buildings. In addition, since architecture features commonly appear symmetrically, while some of the buildings lack symmetric views because of their design or space constraints, the dataset is doubled by mirroring all the images. Since the dataset contains tourist photos from various angles and shooting compositions, there is no need to perform extra transform-based data augmentation.

Art Deco (17/760)	**Art Nouveau (13/746)**	**Baroque (21/3306)**	**Bauhaus & International (15/1052)**
Basilica of the Sacred Heart Brussels	Casa Amatller	Blenheim Palace	Alan I W Frank House
Brooklyn Central Library	Casa Batllo	Cathedral of Santiago de Compostela	Bauhaus Archive
Chicago Board of Trade Building	Casa Mila	Charlottenburg Palace	Bauhaus Building Dessau
Chrysler Building	Cat House Riga	Church of the Gesu	Bauhaus Museum Dessau
Daily Express Building London	Eliseyev Emporium	Drottningholm Palace	Bauhaus Museum Weimar
Eastern Columbia Building	Gresham Palace	Karlskirche	Fagus Factory
Guardian Building	Hotel Metropol Moscow	Les Invalides	Glaspaleis
Hoover Building	House with Chimaeras	Melk Abbey	Haus am Horn
Kennedy-Warren Apartment Building	Jubilee Synagogue	Mexico City Metropolitan Cathedral	Masters' Houses Walter Gropius
Marine Air Terminal	Obecni Dum	Nymphenburg Palace	Max Liebling House
Mutual Heights	Palacio de Bellas Artes	Palace of Versailles	McGraw Hill Building
Palacio de Bellas Artes	Sagrada Familia	Piazza Navona	Sonneveld House
Perelman Building	Secession Building	Royal Palace of Caserta	Villa Stein
Teatro Eden		Royal Palace of Madrid	Villa Tugendhat
The Kean Building		Santa Maria MaggioreSantiago de Compostela Cathedral	Weizmann House
United States Custom House Philadelphia		Schönbrunn Palace	
Vanity Ballroom Building		St. Paul's Cathedral	
		St. Peter's Basilica	
		Winter Palace	
		Zwinger Palace	

Brutalist (17/1034)	**Byzantine (14/1276)**	**Chicago School (16/676)**	**Chinese (16/2778)**
Bank of London and South America	Basilica Cistern	Auditorium Building	Foguang Temple Great East Hall (佛光寺东大殿)
Barbican Centre and Estate	Basilica of San Vitale	Chicago Savings Bank Building	Liuhe Pagoda (六和塔)
Boston City Hall	Basilica of Sant'Apollinare Classe	Guaranty Building	Huayan Temple (华严寺)
Buffalo City Court Building	Basilica of Sant'apollinare Nuovo	Home Insurance Building	Nanchan Temple (南禅寺)
Cite Radieuse	Chora Church	Ludington Building	Giant Wild Goose Pagoda (大雁塔)
Geisel Library	Church of Holy Apostles	Manhattan Apartments	Tianyi Pavilion (天一阁)
Habitat 67	Church of Saint Catherine Thessaloniki	Marquette Building	Fengguo Temple (奉国寺)
National Assembly Building Bangladesh	Daphni Monastery	Marshall Field and Company Building	Temple of Confucius (孔庙)
Paul Rudolph Hall	Hagia Irene	Monadnock Building	Yueyang Tower (岳阳楼)
Royal National Theatre	Hagia Sophia Istanbul	Railway Exchange Building	Chengde Mountain Resort (承德避暑山庄)
SESC Pompeia	Hagia Sophia Thessaloniki	Redmont Hotel	The Palace Musem (故宫)
The Breuer Building	Hosios Loukas	Reliance Building	Jinci (晋祠)
The Met Breuer	Our Lady of Saidnaya Monastery	Rookery Building	Dule Temple (独乐寺)
Trellick Tower	Saint Mark's Basilica	Sullivan Center	White Horse Temple (白马寺)
Tripleone Somerset		Tacoma Building	Mukden Palace (盛京宫殿)
Western City Gate		Toof Building	Bell Tower of Xi'an (西安钟楼)
Wotruba Church			
Deconstructivism (18/1608)	**Gothic (20/2010)**	**Modernism (15/1356)**	**Neoclassical & Beaux-arts (20/1468)**
8 Spruce Street	Amiens Cathedral	Dallas City Hall	Kenwood House
Guggenheim Museum Bilbao	Chartres Cathedral	Palacio da Alvorada	Altes Museum
Seattle Central Library	Lincoln Cathedral	Fallingwater	New York Public Library Main Branch
Antwerp Port Office	Salisbury Cathedral	Phillips Exeter Academy Library	Buckingham Palace
Imperial War Museum Manchester	Basilica of Saint-Denis	Guggenheim Museum NYC	Palais Garnier
UFA cinema center	Cologne Cathedral	S. R. Crown Hall	Chiswick House
Art Gallery of Ontario	Notre-Dame de Paris	Isokon Building London	Pantheon Paris
Jewish Museum Berlin	Seville Cathedral	Sainte Marie de La Tourette	Eglise de la Madeleine
Vitra Design Museum	Brussels Town Hall	John F. Kennedy Presidential Library and Museum	Petit Palais
Beijing National Stadium	Florence Cathedral	Salk Institute for Biological Studies	Egyptian Museum
London Aquatics Centre	Notre-Dame de Reims	National Gallery East Building	San Francisco City Hall
Vitra fire station	St. Stephen's Cathedral	Seagram Building	El Capitolio
CCTV Headquarters	Canterbury Cathedral	Neue National Galerie	Somerset House
Lou Ruvo center for brain health	Jerónimos Monastery	Villa Savoye	General Grant National Memorial
Walt Disney Concert Hall	Orvieto Cathedral	Notre Dame du Haut	Surrogate's Courthouse
Dancing House Prague	Strasbourg Cathedral		Hotel Indigo Dallas Downtown
Museum of Pop Culture	Cathedral Church of Milan		Thomas Jefferson Building
Wexner Centre for the Arts	King's College Chapel		United States Capitol
	Saint Vitus Cathedral		Jefferson Memorial
	Westminster Abbey		United States Post Office and Courthouse Texarkana
Palladian (14/643)	**Postmodernism (18/1370)**	**Romanesque (18/1402)**	**Russian Revival (13/2436)**
Castletown House	400 West Market	Aachen Cathedral	Alexander Nevsky Cathedral
Parliament House Dublin	James R. Thompson Center	Cathedral of Monreale	Kronstadt Naval Cathedral
Villa Rocca Pisani	Neue Staatsgalerie	Lund Cathedral	Sobor Aleksandra Nevskogo
Chiswick House	TC Energy Center	Speyer Cathedral	Cathedral of Christ the Saviour
Queen's House Greenwich	550 Madison Avenue	Autun Cathedral	Moscow Rizhsky railway station
Villa la Rotonda	Kindergarten Wolfartsweier	Cefalù Cathedral	St Nicholas Russian Orthodox Cathedral
Holkham Hall	PPG Place Pittsburgh	Maria Laach Abbey	GUM department store

Stourhead House	Team Disney Building	Tower of London	Our Lady of Iveron Cathedral
White House	Binoculars Building	Basilica of Saint-Sernin	St. Michael's Cathedral Izhevsk
Kedleston Hall	Kyoto Concert Hall	Charlemagne's Chapel	Grand Kremlin Palace
The Rotunda University of Virginia	Piazza d'Italia	Parma Cathedral	Puhtitsa Convent
Woburn Abbey	Vanna Venturi House	Trier Cathedral	Uspenski Cathedral
Leinster House	Bonnefantenmuseum	Basilica of Sant'Ambrogio	Savior on the Spilled Blood
The Royal Crescent	Les Orgues de Flandre	Church of the Holy Sepulchre	
	Portland Building	Pisa Cathedral	
	Denver Public Library	Ca' Loredan	
	Lipstick Building	Durham Cathedral	
	SIS Building	San Miniato al Monte	

Table 1: Architectures chosen for each style. The numbers in brackets denote number of architectures / number of photos

3.4 Supporting Components

The backend will be a web framework that supports Html requests and calls to the above models, including feeding the input into classification module, transfer module and getting the results as the response.

The frontend will be an interactive webpage that communicates with the backend for the user.

4. Implementation

4.1 Implementation list

Classification: ResNet50, using Python and fastai vision

Depth prediction: Affine-invariant depth prediction model, using Python and PyTorch

Image-to-image translation: Pix2pix conditional generative adversarial network, using Python and PyTorch

Frontend: Interactive webpage, using html, css and JavaScript

Backend: Web application, using Flask

4.2 Pretrained model list

Classification: ResNet50
Trained by the dataset mentioned above, 12 epochs

Depth prediction: Affine-invariant depth prediction model
Pretrained by the author

Image-to-image translation: Pix2pix
Base model, trained by all 24720 images, 30 epochs
16 copies of transfer learning models, trained by images of corresponding styles, ≈120000 iterations (no batch normalization) each

4.3 Sequence diagram

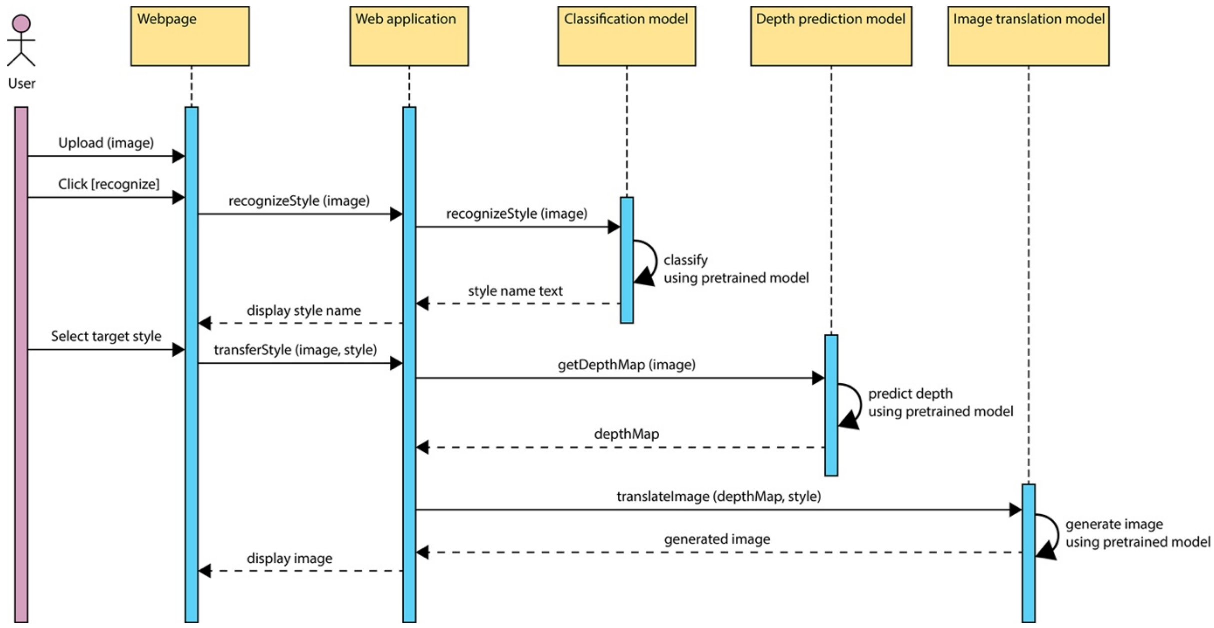

Figure 3: Sequence diagram of this project

5. Testing results

5.1 Classifier testing results

A model is trained with the original 25-class dataset in [4] to compare it with previous works. The result accuracy is 95.25%, much higher than Xu *et al.'s* DPM + MLLR model and close to Zhao *et al.'s* DPM + Feature Extraction + SVM (see Table 2). Thus, ResNet is an efficient solution to architecture classification tasks considering both performance and model complexity. Since the revised dataset has 16 classes, it is reasonable that ResNet achieves a higher accuracy of 97.45% after the 12th epoch (see Table 3).

The confusion matrix (see Table 4) shows that the model can distinguish most of the styles well, except Art Deco and Art Nouveau. This is a good result since both styles are originated from visual art movements instead of independent evolution of architectural design (see Fig. 4), which results in various and mixed appearances.

DPM + MLLR + Spatial Pyramid (Xu et al.)	DPM + Feature Extraction + SVM (Zhao et al.)	ResNet50 (This project)
46.21%	98.57%	95.25%

Table 2: Results trained on Xu *et al*'s 25-class architecture dataset

Epoch	Training loss	Validation loss	Accuracy
0	0.803656	0.504248	0.843998
1	0.568100	0.367579	0.892095
2	0.385303	0.300798	0.908406
3	0.231245	0.222760	0.932246
4	0.135972	0.177396	0.950230
5	0.096788	0.165861	0.951903

6	0.092389	0.151886	0.953158
7	0.098452	0.110531	0.967378
8	0.063302	0.113434	0.968214
9	0.039744	0.102227	0.973233
10	0.024057	0.097746	0.974069
11	0.018166	0.097259	0.974488

Table 3: Accuracy improvement throughout 12 epochs trained on this project's 16-class dataset

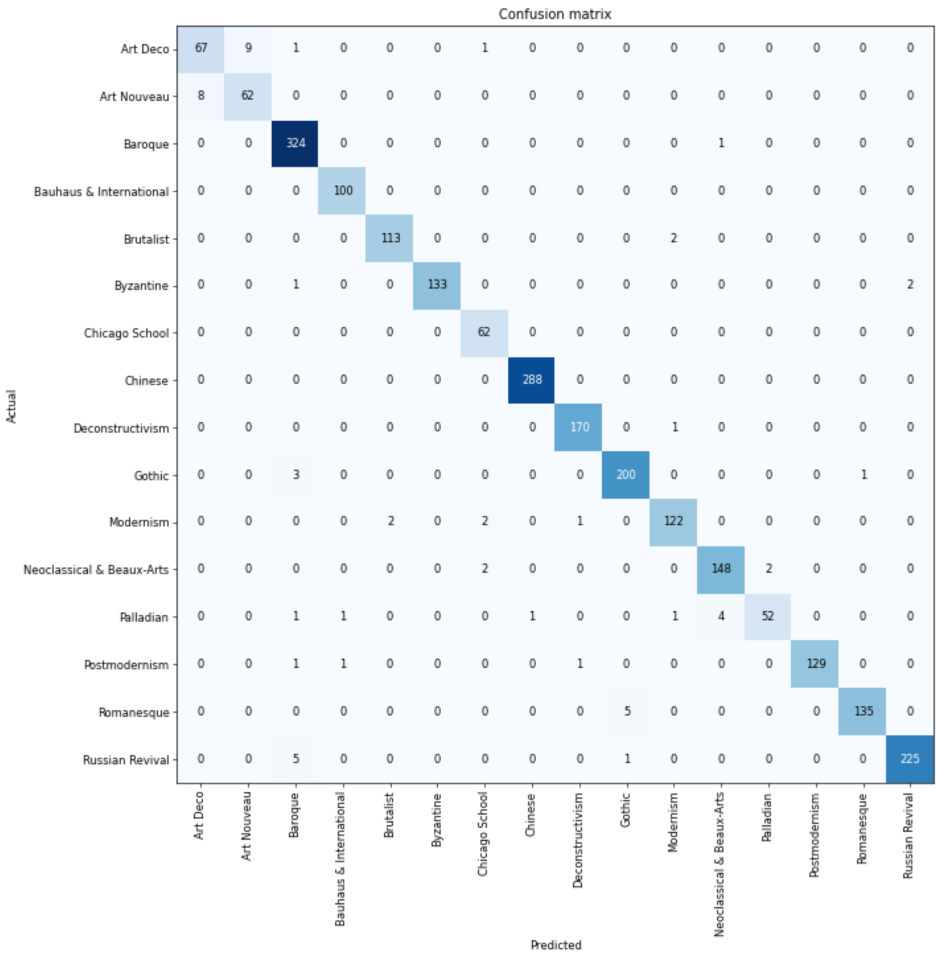

Table 4: Resulted confusion matrix of the 16-class dataset

Basilica of the Sacred Heart (Art Deco) Gresham Palace (Art Nouveau)

Chrysler Building (Art Deco) Sagrada Familia (Art Nouveau)

Figure 4: Examples from Art Deco and Art Nouveau. Both are flexible styles derived from visual art movement instead of design evolution, leading to the result in confusion matrix

5.2 Style transfer testing results

5.2.1 Comparison between different methods

A total of four methods are tested:
1) Neural style transfer with Artflow+WCT model
2) cGAN image translation with Pix2pix model and condition images using rectangular semantic segmentation method addressed in [21]
3) Pix2pix model with condition images using depth map, trained without transfer learning
4) Pix2pix model with condition images using depth map, trained with transfer learning

The testing task is to transfer Neoclassical style onto the famous modernist house Villa Savoye. (see Fig. 5)

For 1), a random Neoclassical building from the dataset and Villa Savoye are input as to style and content images. The result from Artflow + WCT pre-trained model shows little change either in general shape or in the detailed component of the building. Instead, a large portion of blue from the sky in style image is transferred onto the content image like a filter.

For 2), the result suffers from the rigid shape of detailed components, which leads to poor performance, especially when transferring between styles with significant differences. Villa Savoye is a representative example: extremely wide windows and thin pillars do not exist before the industrial revolution and modernism movement due to technical constraints. Thus, forcing the algorithm to apply traditional styles on revolutionary component dimensions results in a highly unrealistic image.

For 3), compared to the previous method, it successfully captures the shape of the building while completely ignoring the detailed components. Neoclassical style trained only by 1468 photos of Neoclassical buildings is transferred onto Villa Savoye, showing recognizable components like windows and pillars. However, the result is far from realistic.

For 4), the quality is better than 3). The main target is much smoother, the pillars are more recognizable, the shaded area is clearer, and the surrounding environment is more realistic.

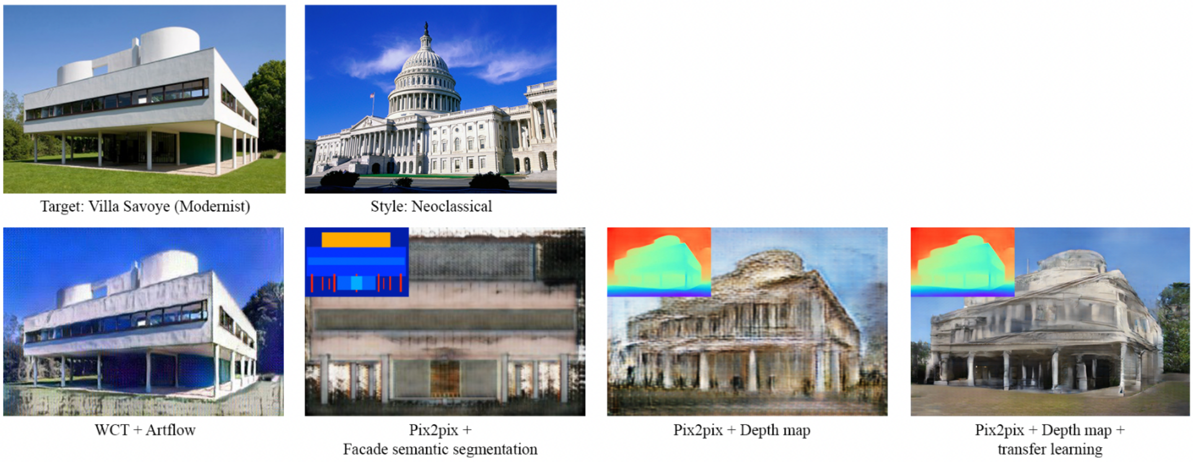

Figure 5: Result of transferring neoclassical style onto Villa Savoye

5.2.2 Comparison of different transfer directions

Eight architectures for eight styles are selected from 16 styles to test transferring performance between each other. A matrix is built with the 64 results (see Fig. 6). The result identifies various problems in the current stage.

1) The results on the diagonal line should remain the same as the original photo. 4 of them are achieved well, while the rest become unrealistic. This may be caused by an unbalanced number of photos for each building within one style. For example, the dataset of Art Deco style has 760 photos, but only 36 of them belong to Chrysler Building used in this matrix. Worse, Chrysler Building is the only skyscraper in Art Deco style. As a result, the model mainly learned from low Art Deco buildings like churches and libraries and failed to transfer their features onto the skyscraper (see Fig. 7).

2) Some of the results overfit their original styles. For example, the cylinder-like component of the Postmodernist Team Disney Building almost remains the same across different results. Only the Russian Revival style is successfully applied on it. This may be caused by the constraints on shapes brought by the depth map. There is no cylinder-like component appearing in other styles. Therefore, the model did not learn anything in the transfer learning step and simply applied the original style features learned from the base model learning step (see Fig. 8).

3) The similar problem caused by shape also appears elsewhere, especially in Chinese style. Chinese architectures have huge eaves, the rough shape of which are captured by the depth map. However, such eaves do not exist in other styles. As a result, transferring the Chinese style to other architectures or reversing directions results in a messy look (see Fig. 9).

Figure 6: Transfer matrix of selected 8 styles

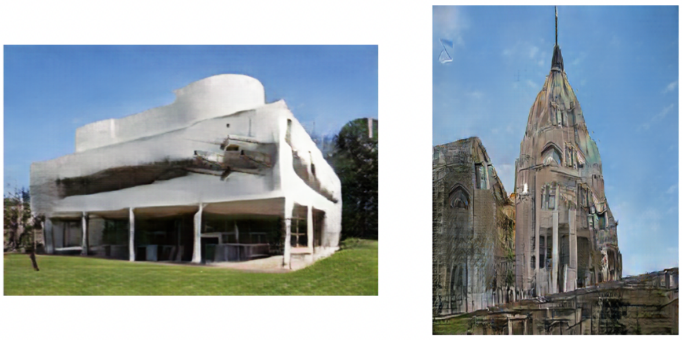

Figure 7: Two results on the diagonal line. Transferring Modernist style onto modernist architecture looks realistic, while Art Deco style performs bad.

Figure 8: Two results of Postmodernist Team Disney Building. Transferring Neoclassical style on to it fails to apply noticeable change on its cylinder-like component, while transferring Russian Revival style applies color from the top of Russian cathedrals.

Figure 9: Style exchanging of Chinese style with both Gothic and Modernist styles results in failure

6. Conclusion

6.1 Critical review

6.1.1 Selection and Change of Direction

I chose architectural style as my project topic since I am really interested in it. However, I did not make a good study of feasibility of this topic, which led to mismatches between the project plan and the final report. Plans including design models with higher performance are abandoned in the middle stage.

6.1.2 Dataset

Based on my additional studies in architecture, I believe that my choices of styles are more reasonable. However, considering the complicated relationship between various styles, I think the dataset still needs further and more professional consideration.

6.1.3 Classification

I have to admit that my literature review of this problem is insufficient, which caused a sudden change in the model to implement at a very late stage. The problem is that I limited my literature review to just the classification of architecture and completely ignored the possibility of bringing better classification models elsewhere to this problem. I struggled with understanding and implementing the complicated systems designed by Xu *et al.* [4] and Zhao *et al.* [5], the performance of which has obviously been out of date. Fortunately, I noticed ResNet classification in the end and successfully trained a high accuracy model in time.

6.1.4 Style Transfer

The choice of GAN instead of neural style transfer is a good decision, since in the early stage of testing, I noticed the uniqueness of architectural style relative to painting styles and the flexibility of GAN models. Also, transfer learning is the right direction since architectures from the different eras still share large amount of common features. However, in the middle stage, I realized that while GANs are powerful tools, researching and training them require large amount of time and knowledge. Thus, I struggled a lot with my time management, a large portion of which is occupied by GAN training. Also, a hurry learning from online courses is far from enough to analyze the GAN model's result thoroughly. Thus, in the end, the training of GAN still looks insufficient, creating highly artificial images in some particular translation direction. Also, as mentioned above, the depth map preprocessing method I came up with has its large weakness, but in the end, I do not have time to tackle it or come up with alternative methods.

6.2 Summary of Achievements

Overall, although the quality of photo synthesis is not high enough, the project has successfully constructed the desired system. In fact, I have already created an experimental design work using the middle-stage model (see Fig. 10). The system and my work have attracted attention from professionals from the field of architecture, varying from a local design biennale to offers from department of architecture of universities including MIT and Harvard.

Figure 10: My experimental design work *Automonument* with help of synthesized images from this project

6.2.1 Classification

ResNet has achieved an excellent performance on this classification task, better than the performance of the systems mentioned in the literature review. Also, ResNet is a popular network easier to implement on a wide range of platforms.

6.2.2 Style Transfer

My depth map and transfer learning method has reached visually acceptable results in some easy translation directions. Also, it shows enough variety and solid potential of generalization to a broader range of problems and even more creative usage since depth map is easy to generate and even synthesize.

6.3 Possible Extensions

Overall, it is obvious that there is a lot of space for improvement in this project, both in terms of performance and application range. First, the depth map is not the only choice for the condition image. Semantic segmentation could still be worth trying despite decreasing the potential variety by the too precise label of objects and building components. Second, CycleGAN-like unpaired image-to-image translation is also a potential solution to this task. As CycleGAN is free from intermediate condition images, the result could be more natural in terms of some detailed components of the buildings. Last, this project could be expanded to video processing, which may produce an interesting result in terms of artistic inspirations.

6. Monthly Log

6.1 October

Take further survey in literature review alongside EN4262 assignments and have gained more valuable resources.
Started learning about basics of machine learning development, and survey of image dataset.

6.2 November

Some efforts on EN4262 Presentation
Tested style transfer method (using Artflow), doesn't work well.
Tested cGAN method (using Pix2pix), the result is profound for building with similar structures, but bad for styles with huge differences (The result of generating modernism house with trained model of European houses is unrecogizable)
Currently looking for better ways of generating CMP-Facade-like semantic segmentation

6.3 December

Much better result generated using depth map as the labeling method, currently using DiverseDepth's depth prediction. But still expect a better method to describe shape, since depth prediction produce inconsistent result due to angle, distance, etc.
More work have done and still needed on preparation of dataset for cGAN training with new labeling method.

6.4 January

More work have done on training and testing depth map method for cGAN model.
Additional effort in learning implementation of DCNN and cGAN.

6.5 February

Collecting more dataset
Setting up frontend
Investigating possibilities of model modification

6.6 March

Finish implementing interactive webpage and final report

7. References

[1] G. Kubler, "Toward a reductive theory of visual style," in The Concept of Style, 1st ed. B. Lang, Ed. Ithaca and London: Cornell University Press, 1987, ch. 5, pp. 163-173.

[2] M. Fletcher and R. Polley, Architectural Styles: A Visual Guide, 1st ed. Princeton: Princeton University Press, 2020, pp. 6-11.

[3] Alfred Dwight Foster Hamlin, History Architectural Styles, 1st ed. Sacramento: Creative Media Partners, LLC, 2018.

[4] A. Goel, M. Juneja, and C. Jawahar, “Are buildings only instances?” in Proceedings of the Eighth Indian Conference on computer vision, graphics and image processing, 2012, pp. 1–8, doi: 10.1145/2425333.2425334.

[5] Z. Xu, D. Tao, Y. Zhang, J. Wu, and A. C. Tsoi, “Architectural Style Classification Using Multinomial Latent Logistic Regression,” in Computer Vision – ECCV 2014, pp. 600–615, doi: 10.1007/978-3-319-10590-1_39.

[6] P. Zhao, Q. Miao, J. Song, Y. Qi, R. Liu, and D. Ge, “Architectural Style Classification Based on Feature Extraction Module,” IEEE access, vol. 6, pp. 52598–52606, 2018, doi: 10.1109/ACCESS.2018.2869976.

[7] Y. Yoshimura, B. Cai, Z. Wang, and C. Ratti, “Deep Learning Architect: Classification for Architectural Design through the Eye of Artificial Intelligence,” 2018.

[8] “Architecture time warp: 7 iconic buildings reimagined,” May 5, 2018. Accessed on: Dec. 3, 2021. [Online]. Available: https://www.nbcnews.com/mach/tech/time-warp-architecture-seven-iconic-buildings-reimagined-ncsl871061/engineering-triumph-forged-nation-panama-canal-turns-100-n181211

[9] L. A. Gatys, A. S. Ecker, and M. Bethge, “A Neural Algorithm of Artistic Style,” 2015.

[10] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image Style Transfer Using Convolutional Neural Networks," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2414–2423, doi: 10.1109/CVPR.2016.265.

[11] L. A. Gatys, A. S. Ecker, M. Bethge, A. Hertzmann, and E. Shechtman, "Controlling Perceptual Factors in Neural Style Transfer," In CVPR, 2016.

[12] X. Huang and S. Belongie, "Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 1510–1519, doi: 10.1109/ICCV.2017.167.

[13] P. Chandran, G. Zoss, P. Gotardo, M. Gross, D. Bradley, "Adaptive Convolutions for Structure-Aware Style Transfer," in 2021 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

[14] Y.n Li, C. Fang, J. Yang, Z. Wang, X. Lu, M. Yang. Universal style transfer via feature transforms. In Advances in Neural Information Processing Systems, pages 386–396, 2017.

[15] J. Yoo, Y. Uh, S. Chun, B. Kang, and J.-W. Ha, "Photorealistic Style Transfer via Wavelet Transforms," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 9035–9044, doi: 10.1109/ICCV.2019.00913.

[16] L. Sheng, Z. Lin, J. Shao, and X. Wang. Avatar-net: multi-scale zero-shot style transfer by feature decoration. In CVPR, 2018.

[17] J. An, S. Huang, Y. Song, D. Dou, W. Liu, J. Luo, "ArtFlow: Unbiased Image Style Transfer via Reversible Neural Flows," in 2021 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

[18] X. Wang and A. Gupta, "Generative Image Modeling Using Style and Structure Adversarial Networks," in Computer Vision – ECCV 2016, 2016, pp. 318–335, doi: 10.1007/978-3-319-46493-0_20.

[19] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[20] K.-H. Liu, T.-J. Liu, C.-C. Wang, H.-H. Liu, and S.-C. Pei, "Modern Architecture Style Transfer for Ruin or Old Buildings," in 2019 IEEE International Symposium on Circuits and Systems (ISCAS), 2019, pp. 1–5, doi: 10.1109/ISCAS.2019.8702121.

[21] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," arXiv preprint arXiv:1703.10593, 2017.

[22] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in International Conference on Learning Representations (ICLR), 2015.

[23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[24] Refik Anadol Studio, “Machine Hallucinations.” New York: Artechouse NYC, 2019.

[25] Maxime Lefebvre, “Periaktoi: An A.I. interpretation of the Scaenae Frons.” [Online Video]. Available: https://youtu.be/IhiU7rgLOmw

[26] R. Tyleček and R. Šára, “Spatial Pattern Templates for Recognition of Objects with Regular Structure,” in Pattern Recognition, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 364–374.

[27] Wei Yin, Xinlong Wang, Chunhua Shen, Yifan Liu, Zhi Tian, Songcen Xu, Changming Sun, and Dou Renyin, “DiverseDepth: Affine-invariant Depth Prediction Using Diverse Data,” in 2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020.