BIM Style Restoration Based on Image Retrieval and Object Location Using Convolutional Neural Network

Yalong Yang 1,2,3, Yuanhang Wang 1,2,3, Xiaoping Zhou 4, Liangliang Su 1,2,3,* and Qizhi Hu 1,2,3

1 Anhui Province Key Laboratory of Intelligent Building and Building Energy Saving, Anhui Jianzhu University, Hefei 230022, China
2 Anhui Institute of Strategic Study on Carbon Dioxide Emissions Peak and Carbon Neutrality in Urban-Rural Development, Hefei 230022, China
3 School of Electronic and Information Engineering, Anhui Jianzhu University, Hefei 230601, China
4 Beijing Key Laboratory of Intelligent Processing for Building Big Data, Beijing University of Civil Engineering and Architecture, Beijing 100044, China
* Correspondence: llsu_yz@ahjzu.edu.cn

Abstract: BIM is one of the main technical ways to realize building informatization, and the model's texture is essential to its style design during BIM construction. However, the texture maps provided by mainstream BIM software are not realistic enough and monotonous to meet the actual needs of users for the model style. Therefore, an interior furniture BIM style restoration method was proposed based on image retrieval and object location using convolutional neural network. First, two types of furniture images, namely grayscale contour images from BIM software and real images from the Internet, were collected to train the following network model. Second, a multi-feature weighted fusion neural network model based on an attention mechanism (AM-rVGG) was proposed, which focused on the structural information of furniture images to retrieve the most similar real image, and then some furniture image patches from the retrieved one were generated with object location and random cropping techniques as the candidate texture maps of the furniture BIM. Finally, the candidate ones were fed back into the BIM software to realize the restoration of the furniture BIM style. The experimental results showed that the average retrieval accuracy of the proposed network model was 83.1%, and the obtained texture maps could effectively restore the real style of the furniture BIM. This work provides a new idea for restoring the realism in other BIM.

Keywords: convolutional neural network; image retrieval; object location; BIM; texture map

1. Introduction

Building digitalization has become an inevitable trend of transformation and upgrading in construction in recent years. As one of the most effective technologies to realize building informatization, building information modeling (BIM) can digitally express the building facilities’ physical and functional characteristics. It also provides reliable shared information resources for all kinds of decision-making over the whole life cycle of buildings [1,2]. Therefore, the authenticity and integrity of the BIM’s description of building facilities is an important basis for determining whether the model is available. As one of the essential attributes reflecting its style, the texture of the BIM plays a vital role in the quality of the model, especially in the BIM of architectural cultural heritage [3–5].

Currently, BIM mainly comes from modeling software and various network platforms, but the model visualization effect provided is challenging in terms of meeting the authenticity requirements. On one hand, the texture map provided by the modeling software is too singular and lack diversity compared to the actual needs. This is different to building a model that is visualized with a natural appearance such as Autodesk Revit. On the other hand, due to the personalized style design requirements, the model provided by the network platform needs to be reconstructed, and it still lacks the natural texture map.
At the same time, due to the influence of human subjectivity, there is a material (texture) imbalance between the model texture style and the actual component [6–8]. In sum, it is very important for the style restoration of BIM to obtain a texture map of the real object that is highly similar to the model.

Cross-domain image retrieval offers a feasible way to obtain real texture in BIM. That is, given a grayscale contour image of the BIM from BIM software, similar natural images are retrieved from the Internet [9]. Then, the nature images will produce the same small image blocks as candidate texture maps through object location and random clipping techniques. Generally, the higher the similarity between the retrieved natural images and the BIM image, the more beneficial the obtained texture maps are to restore the true style of the BIM.

2. Related Work

Currently, deep neural network models have achieved a series of remarkable results in the field of computer vision such as AlexNet [10], VGGNet [11], ResNet [12], and so on. In recent years, with the extension of deep learning technology, scholars have carried out some related work in BIM, mainly involving the BIM building component, for instance, classification [13–15] and architectural style classification [16–19]. However, as far as we know, there have been few reports on the use of deep neural networks to obtain the texture maps needed by BIM from real scenes and to restore their styles. In addition, the BIM includes the local component style and the overall model style. At the same time, in the real world, all kinds of furniture have different styles because of their different appearance characteristics and play an essential role in the interior home decoration style [18]. Therefore, this paper took the furniture image as the research object.

For furniture images with complex backgrounds and high similarity between classes, scholars have put forward relevant research. For example, in 2016, Bermeitinger et al. optimized the classification performance of neoclassical furniture images by image enhancement through the VGGNet16 model [19]. In 2017, Hu et al. classified furniture styles through manual features, depth features, or their combination, and the results showed that the combination of depth features and manual features had a better classification effect [20]. Hu et al. used the VGGNet model to classify the styles of three types of architectural pictures and a variety of non-architectural objects, and verified the advantages of a deep neural network in style cognition [21]. In 2021, Du et al. took furniture style classification as the research goal, processed the deep features of VGGNet16 by Gram transformation, and achieved good results in furniture style classification [22]. Although the above four methods optimized the classification performance of furniture from different perspectives, they did not optimize the backbone network. In 2018, Wang et al. proposed an AlexNet-S network model combined with an image similarity measurement algorithm to remove duplicate and irrelevant samples in the furniture image database [23]. Then, Luo Xia et al. classified furniture images through AlexNet based on feature fusion, but the above two methods were not further verified on other convolutional neural networks [24]. Sui et al. imitated the human attention mechanism and proposed a convolutional neural network (CNN) model that highlighted the color, contour, and other information of furniture images that makes up for the traditional CNN. The deficiencies in the image color and other features were ignored, and there was a lack of attention to the image channel field [25]. Ataer et al. simplified the VGGNet model by removing the fully connected layer to classify different styles of interior design renderings. Although this method could effectively retain the network convolution features, there was no further study on furniture classification in the renderings [26]. In the research of BIM in 2022, Zhou et al. used the YOLO neural network to detect the position and size information of the object from the camera video. They realized the effect of real-time restoration of the BIM according to the actual scene [27].

Based on the above analysis, in order to restore the style of furniture BIM, a multi-feature weighted fusion neural network model based on attention mechanism (AM-rVGG) was first proposed. The model focuses on the furniture image features from the spatial
and channel domains of the image, and balances the convolutional features of different depths through multi-feature weighted fusion to improve the accuracy of the cross-domain retrieval of the neural network, and then retrieve the real furniture image with high similarity to the furniture BIM. Second, a texture map generation method is proposed to obtain the texture map of the real furniture image. Finally, the obtained texture map is fed back to the BIM to restore the style of the BIM.

3. Method

For the problem of BIM style restoration, in order to obtain a realistic furniture image that is most similar to the BIM, first, the contour map of the furniture model is obtained. At the same time, the real furniture image dataset is established by the image preprocessing operation. Then, the contour image of the furniture model is input into the AM-rVGG network model, and the real furniture image similar to the contour image of the furniture model is retrieved in the real furniture image dataset. To obtain the candidate texture map of the BIM, the real furniture image is processed by the texture map restoration method. The obtained texture map is finally fed back to the BIM to realize the style restoration of the BIM. The general framework of the method is shown in Figure 1.

![Figure 1](image_url)

**Figure 1.** The overall framework of the BIM style restoration method.

3.1. AM-rVGG Network Model

VGGNet [11] was first proposed in the ImageNet image classification competition in 2014. It mainly has four structures with different depths, among which VGGNet16 is the most widely used. The VGGNet16 network can be divided into five convolutions and one full connection. Each convolution consists of two or three convolution layers. Each convolution layer uses a $3 \times 3$ convolution kernel, and the Relu activation function is used after the convolution layer. At the same time, each convolution layer is pooled by a $2 \times 2$ maximum pooling layer.

The VGGNet16 model has been widely used in image classification with deep network depth, but with the increase in the number of convolution layers, the image detail
information will be easily lost. In this paper, the attention mechanism was introduced into the VGGNet16 network model, and the method of multi-feature weighted fusion was used to compensate for the network details. The structure of the network model is shown in Figure 2.

![Figure 2. AM-rVGG model framework.](image)

According to the AM-rVGG model framework in Figure 2, for furniture images with complex backgrounds, the high similarity between classes and large differences within categories, details such as texture and shape, which can be used to distinguish classes, are not obvious enough. Moreover, when the image is processed by the convolutional layer and the fully connected layer to obtain high-level semantic features, it will cause further loss of detailed information. In order to effectively retain the detailed information of the bottom layer of convolution and balance the features of different convolution depths, a multi-feature fusion method is proposed. First, the convolution layer of the VGGNet16 is divided into five convolutions, and the convolution features of different depths are selectively fused. Specifically, the fourth and seventh convolutional output features of VGGNet16 are fused, respectively, and the convolution output feature of the tenth layer and the convolution output feature of the thirteenth layer are fused. For the network model, the features of the last fully connected layer are input as the final classification features. In order to make better use of the fused features for classification, the features of the last convolutional layer and the fully connected layer are spliced and fused.

Multi-feature fusion can retain the convolution features of different depths without weight and then balance the features of different depths. However, in order to make the network model have a more obvious image classification ability, it is necessary to fuse the detailed information that can represent the image category. Therefore, the attention mechanism (convolutional block attention module, CBAM) can be used to highlight the weights of the detail features to improve the classification performance of the network [28]. In the CBAM module, the input feature is used to infer the attention map along two independent dimensions (channel and space) in turn and then multiplied with the input feature map to generate the recalibrated feature. Assuming an input feature as a one-dimensional channel attention feature descriptor and as a two-dimensional spatial feature descriptor, the overall process of the attention mechanism is:

$$
\begin{align*}
F' &= M_c(F) \otimes F \\
F'' &= M_s(F') \otimes F'
\end{align*}
$$

(1)
where $\otimes$ stands for the multiplication of elements. In order to fuse more representative features and avoid the loss of information due to the attention module, a weighted feature fusion method based on the attention mechanism is proposed. In this method, the re-weighted features of the attention module are further fused with the input features, and the fused features are input to the next convolution. The feature weighted fusion method based on the attention mechanism is shown in Figure 3.

![Feature weighted fusion method based on the attention mechanism](image)

Figure 3. Feature weighted fusion method based on the attention mechanism.

More comprehensive and typical features can be obtained through the multi-feature weighted fusion network based on the attention mechanism. Before the features are input into the fully connected layer, they need to be processed in the channel directly through the flattening operation. The traditional VGGNet16 model has three fully connected layers, but only the features before the last fully connected layer are used for classification. The features before the fully connected layer may cause further loss of feature information during the flattening operation and the fully connected layer; therefore, based on the traditional VGGNet16 model, the second layer of full connection is removed, and the features weighted based on the attention mechanism are spliced with the last fully connected feature, which is used for the final classification.

3.2. Texture Map Restore

The furniture image finally obtained by the above method often has a complex background, and the texture image of the furniture itself is not uniform. In order to generate the texture map of the furniture as accurately as possible, the complex background of the furniture image should first be filtered. For first-stage target detection algorithms such as YOLO and SSD, they can quickly identify and select targets from complex image backgrounds [29,30]. Therefore, in this paper, the object detection algorithm YOLO was used to obtain the minimum block diagram of the real furniture object [29]. The image was cropped according to the generated block diagram coordinates, and then the individual furniture image was cropped from the complex original image. However, the individual furniture image contained the textures of different structures of the furniture itself, in order to obtain the texture map of each part to restore the BIM style. Therefore, a texture map restore method was proposed. First, the furniture images obtained by clipping were randomly clipped according to the set clipping size and quantity. Then, the average pixels of the images obtained by clipping were counted in different regions. Finally, the images with large differences in the average pixels of the regions were removed. The other images were saved according to the average pixels of different regions, and the images were spliced and synthesized. Finally, the texture maps of different structures of furniture were obtained. The texture map restoration method framework is shown in Figure 4.
4. Experimental Results and Analysis

4.1. Dataset

As far as we know, there is no public image dataset for furniture BIM retrieval. Therefore, a separate dataset was constructed that included two types of images, namely, real furniture images and furniture model contours. Real furniture images through the web crawler technology, according to the category of furniture from Baidu images, downloaded a total of 26,806 real furniture images. Still, the Baidu image platform has the problem of repeated downloading of images; in order to improve the phenomenon of the same image, this paper will randomly retain the real furniture images obtained from the Baidu image platform and finally obtained a total of 11,753 images. These included chairs, beds, tables, windows, and doors. There were one to four sub-categories under each category of real furniture images, and the specific sub-categories are shown in Table 1.

| Major Categories | Subclass (Quantity) |
|------------------|---------------------|
| Chair            | Office chair (1499), dining chair (1667), Chinese chair (806) |
| Bed              | Chinese bed (1154), European bed (1004) |
| Table            | Chinese table (882), European table (889) |
| Window           | Chinese window (674) |
| Door             | Primary and secondary doors (1354), interior doors (1138), European doors (340), Chinese doors (346) |

For the outline drawing of the furniture model, the furniture model was downloaded from Revit and various platforms according to the above classification, and 1035 pictures were captured from different angles. The specific number is shown in Table 2.
Table 2. Breakdown of the major and minor categories of the BIM family images.

| Major Categories | Subclass (Quantity)                                      |
|------------------|--------------------------------------------------------|
| Chair            | Office chair (216), dining chair (129), Chinese chair (96) |
| Bed              | Chinese bed (63), European bed (57)                   |
| Table            | Chinese table (57), European table (45)               |
| Window           | Chinese window (39)                                   |
| Door             | Primary and secondary doors (84), interior doors (42), European doors (192), Chinese doors (15) |

The real furniture image itself had color features, but the contour image of the model lacked color features. In order to reduce the difference between the two different domains and improve the retrieval accuracy, Gaussian filtering and edge extraction were used to process the real furniture images. It unified the size of the two datasets as $224 \times 224$. Figure 5 shows an example of some furniture images from the constructed dataset.

4.2. Experimental Setup

The style restoration of the furniture BIM needs to obtain the texture map of the real furniture image, which is similar to the outline structure of the furniture model. In order to carry out the feature extraction and retrieval experiment of the furniture model outline and real furniture image, methods based on HOG, VGGNet16, AM-rVGG, and AM-rVGG + HOG were designed. The HOG features were obtained based on the strategy in [31]; the AM-rVGG + HOG feature is the cascade fusion of the feature vector of the penultimate fully connected layer of the AM-rVGG model and the HOG feature. When retrieving, the cosine distance is used as the similarity measurement method to obtain the desired actual image similar to the outline structure of the furniture model to be tested and then the texture map is obtained.

Accuracy is used as the evaluation index of network model training, and its mathematical description is shown in Equation (2) where $N$ represents the number of images correctly recognized by the network model, and $M$ represents the number of all images.

$$\text{Accuracy} = \frac{N}{M}$$ (2)
The mean average precision (mAP) was used as the evaluation index of the retrieval experiment, and its mathematical description is shown in Equation (3).

\[
\begin{align*}
\text{Precision} &= \frac{TP}{TP + FP} \\
\text{mAP} &= \frac{1}{n} \sum_{k=1}^{n} \text{Precision}_k
\end{align*}
\]

where precision is the retrieval precision; \(n\) is the number of retrieval times; \(TP\) is the number of images that should be retrieved; \(FN\) is the number that should not be retrieved; and \(FP\) is the number that is incorrectly detected.

The experiment in this paper was based on the Pytorch deep learning framework, and the Adm optimizer was used as the gradient descent algorithm. The initial learning rate was set to 0.0002, the number of training iterations was \(\text{Epochs} = 200\), and the batch data size was \(\text{Batch-Size} = 64\). The furniture dataset was divided into a training set and a test set according to a ratio of 9:1. All lab environments were Intel I9-11900K, Nvidia RTX3090, and 128 GB RAM.

4.3. Experimental Results and Analysis

First, the accuracy comparison images of AM-rVGG and VGGNet16 on the dataset are given, as shown in Figure 6. From the accuracy comparison curve, it can be seen that the convergence speed and accuracy of AM-rVGG network training had been significantly improved, and the accuracy of the first training of the AM-rVGG network reached about 45%. The comparison results of the highest accuracy and the average accuracy of the network in 200 iterations are shown in Table 3.

![Figure 6. Comparison curve of the network training accuracy.](image)

| Network Model | Highest Accuracy | Average Accuracy |
|---------------|-----------------|-----------------|
| VGG           | 0.8665          | 0.8091          |
| Net16         | 0.8665          | 0.8091          |
| AM-rVGG       | 0.8845          | 0.8311          |
It can be seen from Table 3 that compared with the VGGNet16 network model, the best accuracy of the AM-rVGG network model was increased by 1.8%, and the average accuracy of the training was increased by 2.2%. The average accuracy of the training had been significantly improved, which further proved that the convergence effect of the network model in the training process had been considerably improved.

In order to verify the effectiveness of this method more comprehensively, this paper designed retrieval experiments based on the HOG, VGGNet16, AM-rVGG network, and AM-rVGG + HOG method for different types of furniture data, and obtained the mAP values for various kinds of furniture. The comparison results are shown in Table 4.

### Table 4. mAP values for the different types of furniture.

| Serial Number | Image Category       | Quantity | HOG   | VGG Net16 | AM_rVGG | AM_rVGG + HOG |
|---------------|----------------------|----------|-------|-----------|----------|---------------|
| 1             | Office chair         | 216      | 0.4582| 0.1279    | 0.1291   | 0.1281        |
| 2             | Dining chair         | 129      | 0.3371| 0.0988    | 0.1072   | 0.1076        |
| 3             | Chinese style chair  | 96       | 0.0404| 0.0692    | 0.0689   | 0.0702        |
| 4             | European bed         | 57       | 0.1278| 0.0858    | 0.0839   | 0.0918        |
| 5             | Chinese bed          | 63       | 0.0337| 0.0988    | 0.1072   | 0.1076        |
| 6             | Chinese window       | 39       | 0.0168| 0.0593    | 0.0598   | 0.0611        |
| 7             | European table       | 45       | 0.0454| 0.0757    | 0.0783   | 0.0791        |
| 8             | Chinese table        | 57       | 0.0227| 0.0763    | 0.0786   | 0.0755        |
| 9             | European style door  | 192      | 0.0117| 0.0298    | 0.0296   | 0.0296        |
| 10            | Interior door        | 42       | 0.0965| 0.0978    | 0.1017   | 0.1082        |
| 11            | European style door  | 15       | 0.0057| 0.0329    | 0.0312   | 0.0350        |

The Table 4 show that among the 12 categories of furniture images, the mAP values of the seven categories of furniture images were the highest by using the AM-rVGG feature and the HOG feature for fusion, and the mAP values of the three types of furniture images retrieved by using the HOG feature were higher and more prominent. By comparing the retrieval method based on HOG with that based on deep learning, it was found that the retrieval method based on deep learning has obvious advantages for small sample furniture images, indicating that deep learning features have more ability to express the image features. By comparing the retrieval experiments of VGGNet16 and AM-rVGG, it was found that the AM-rVGG method had an advantage in eight kinds of furniture image retrieval experiments, which further verified the effectiveness of the proposed method.

Finally, based on the AM-rVGG method, the mainstream BIM software Revit 2020 was used as a platform. The interior door was used as an example to verify the restoration of the furniture BIM style. First, the outline drawing of the BIM of the indoor door was obtained and processed by the AM-rVGG network model, and the cosine distance was used as the similarity evaluation index to sort and output the real furniture images similar to the BIM. The steps to retrieving the outline drawing of the indoor door model are shown in Figure 7.

![Figure 7. Example of the retrieval result of the BIM door model outline drawing.](image-url)
In general, to obtain a real furniture image, the image with the maximum similarity is used as the input to generate a candidate texture map. For an obtained real furniture image, a minimum block diagram of a real furniture object is first obtained by utilizing a object location algorithm [19] and is output by clipping; the main door body is stripped from a complex background image in the step, then the door in the main body image is processed by a texture map restoration method to obtain a candidate texture map. The obtained texture maps of the door panel, the glass, and the door handle are finally fed back to the Revit modeling software to realize the style restoration of the indoor door model. The style restoration steps are shown in Figure 8.

![Real furniture images](image1)

![Texture map of door plate, glass and door handle](image2)

![BIM model style restore](image3)

**Figure 8.** The BIM style restoration example.

An example of the style restoration for a broader type of furniture BIM is shown in Figure 9.

![furniture BIM model](image4)

![Real furniture image retrieval](image5)

![Real furniture texture map acquisition](image6)

![BIM furniture model style restoration](image7)

**Figure 9.** Examples of the furniture BIM style restoration.

The AM-rVGG method was used for the retrieval experiment, and the experimental results were sorted according to the similarity. In this experiment, the style of the furniture BIM was generated based on the furniture image with the largest similarity. It can be seen from the example that for the model with more complex construction such as the European bed, the real furniture image retrieved had more complex and diverse textures, and the candidate textures processed by the texture map restoration method were also more real, so the overall style of the BIM generated by its style was more real. For a model with a simple structure such as an office chair, the retrieved image texture of the office chair is relatively simple, so the overall style effect of the BIM after texture map feedback is not as obvious as that of a European bed.
5. Conclusions

In order to solve the problem that the texture map provided by the mainstream BIM software was not real enough and single, which cannot meet the needs of users for the BIM style design, in this study, taking advantage of the advantages of a convolutional neural network, a style restoration method of indoor furniture BIM based on a convolutional neural network is proposed. First, in order to train the proposed model, a dataset of grayscale contour images from the BIM software and real images from the Internet were constructed. Second, in order to obtain the most similar real images, a multi-feature weighted fusion neural network model based on the attention mechanism (AM-rVGG) was proposed. The model optimized VGGNet16 by the attention mechanism and feature fusion, and the average retrieval accuracy was improved by 2.2%. Finally, a technique based on object localization and random cropping was used to generate some furniture image blocks from the retrieved furniture images as candidate texture maps for furniture BIM. The candidate texture map was fed back to the BIM software to realize the style restoration of furniture BIM.

Experiments showed that the BIM restored by this scheme could restore the real scene more realistically, which provides a new idea for the research of various model styles under the concept of BIM. With the continuous research of the deep learning algorithm in the field of BIM, it will further promote the development of building digitalization. It is worth mentioning that in the face of complex BIM, the method of this study is still insufficient to extract the local complex texture map, which will be the next research direction of this study.

Author Contributions: Conceptualization, Y.Y. and L.S.; Methodology, Y.W., L.S. and Q.H.; Formal analysis, Y.Y., Y.W. and L.S.; Data curation, Y.W.; Writing—original draft preparation, Y.Y. and Y.W.; Supervision, X.Z. and L.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China, grant number 62001004; the Key Provincial Natural Science Research Projects of Colleges and Universities in Anhui Province, grant number KJ2019A0768; the Key Research and Development Program of Anhui Province, grant number 202104a07020017; and the Academic Funding Program for Top-Notch Talents of Disciplines (Majors) in Universities of Anhui Province, grant number gxbjZD2022028, and the Research Project Reserve of Anhui Jianzhu University, grant number 2020XMK04.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Zhang, J.; Long, Y.; Lv, S.; Xiang, Y. BIM-enabled Modular and Industrialized Construction in China. Procedia Eng. 2016, 145, 1456–1461. [CrossRef]
2. Zheng, Z.; Liao, W.; Lin, J.; Zhou, Y.; Zhang, C.; Lu, X. Digital Twin-Based Investigation of a Building Collapse Accident. Adv. Civ. Eng. 2022, 2022, 9568967. [CrossRef]
3. Hull, J.; Ewart, I.J. Conservation data parameters for BIM-enabled heritage asset management. Autom. Constr. 2020, 119, 103333. [CrossRef]
4. Adamopoulos, E.; Rinaudo, F. Close-Range Sensing and Data Fusion for BIM-enabled heritage asset management. Int. J. Archit. Herit. 2021, 15, 1930–1941. [CrossRef]
5. Cogima, C.K.; Paiva, P.V.V.; Dezen-Kempter, E.; Carvalho, M.A.G.; Soibelman, L. The role of knowledge-based information on BIM for built heritage. In Advances in Informatics and Computing in Civil and Construction Engineering; Springer: Cham, Switzerland, 2019; pp. 27–34.
6. Gunn, T.G. The mechanization of design and manufacturing. Sci. Am. 1982, 247, 114–131. [CrossRef]
7. Machete, R.; Falcão, A.P.; Gonçalves, A.B.; Godinho, M.; Bento, R. Development of a Manuoline style object library for heritage BIM. Int. J. Archit. Herit. 2021, 15, 1930–1941. [CrossRef]
8. Qiu, Q.; Zhou, X.; Zhao, J.; Yang, Y.; Tan, S.; Wang, J.; Liu, J.; Liu, H. From sketch BIM to design BIM: An element identification approach using Industry Foundation Classes and object recognition. Build. Environ. 2021, 188, 107423. [CrossRef]
9. Nie, W.; Zhao, Y.; Nie, J.; Liu, A.A.; Zhao, S. CLN: Cross-Domain Learning Network for 2D Image-Based 3D Shape Retrieval. *IEEE Trans. Circuits Syst. Video Technol.* 2021, 32, 992–1005. [CrossRef]

10. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 2012, 25, 1097–1105. [CrossRef]

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

12. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

13. Kim, J.; Song, J.; Lee, J.K. Recognizing and classifying unknown object in BIM using 2D CNN. In Proceedings of the International Conference on Computer-Aided Architectural Design Futures, Daejeon, Republic of Korea, 26–28 June 2019; pp. 47–57.

14. Wan-qi, W.A.N.G.; Bao-rui, M.A.; Qian, L.I.; Wen-long, L.U.; Yu-shen, L.I.U. Clustering of BIM components based on similarity measurement of attributes. *J. Graph.* 2020, 41, 304.

15. Wang, J.; Su, D.; Zhou, X. BIM model similarity calculation method. *J. Graph.* 2020, 41, 624–631.

16. Dautov, E.; Astafeva, N. Convolutional neural network in the classification of architectural styles of buildings. In Proceedings of the 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), St. Petersburg, Moscow, Russia, 26–28 January 2021; pp. 274–277.

17. Zhao, P.; Miao, Q.; Song, J.; Qi, Y.; Liu, R.; Ge, D. Architectural style classification based on feature extraction module. *IEEE Access* 2018, 6, 52598–52606. [CrossRef]

18. Xia, B.; Li, X.; Shi, H.; Chen, S.; Chen, J. Style classification and prediction of residential buildings based on machine learning. *J. Asian Archit. Build. Eng.* 2020, 19, 714–730. [CrossRef]

19. Bermeitinger, B.; Freitas, A.; Donig, S.; Handschu, S. Object classification in images of Neoclassical furniture using Deep Learning. *Int. Workshop Comput. Hist. Data-Driven Humanit.* 2016, 2016, 109–112.

20. Hu, Z.; Wen, Y.; Liu, L.; Jiang, J.; Hong, R.; Wang, M.; Yan, S. Visual classification of furniture styles. *ACM Trans. Intell. Syst. Tech. (TIST)* 2017, 8, 1–20. [CrossRef]

21. Hu, W. The experiment of neural network on the cognition of style. In Proceedings of the 26th CAADRIA Conference, Hong Kong, China, 29 March–1 April 2021; pp. 61–70.

22. Du, X. FISC: Furniture image style classification model based on Gram transformation. In Proceedings of the 2021 3rd International Conference on Advanced Information Science and System (AISS 2021), Sanya, China, 26–28 November 2021; pp. 1–5.

23. Wang, Y.; Gao, W.; Wang, Y. Application of furniture images selection based on neural network. In *AIP Conference Proceedings*; AIP Publishing LLC: Busan, Republic of Korea, 2018; Volume 1967, p. 040016.

24. Luo, X. Research on Convolutional Neural Network Furniture Image Classification Algorithm Based on Feature Fusion. Master’s Thesis, Nanjing University of Posts and Telecommunications, Nanjing, China, 2019. [CrossRef]

25. Ting-Ting, S.; Ke-Yu, Z.; Hui, Z.; Qiao, H. Interest Points guided Convolution Neural Network for Furniture Styles Classification. In Proceedings of the 2019 6th International Conference on Systems and Informatics (ICSAI), Shanghai, China, 2–4 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1302–1307.

26. Ataer-Cansizoglu, E.; Liu, H.; Weiss, T.; Mitra, A.; Dholakia, D.; Choi, J.W.; Wulin, D. Room style estimation for style-aware recommendation. In Proceedings of the 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), San Diego, CA, USA, 9–11 December 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 267–2673.

27. Zhou, X.; Sun, K.; Wang, J.; Zhou, J.; Feng, C.; Yang, Y.; Zhou, W. Computer Vision Enabled Building Digital Twin Using Building Information Model. *IEEE Trans. Ind. Inform.* 2022, early access. [CrossRef]

28. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* 2017, 30. [CrossRef]

29. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

30. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016; pp. 21–37.

31. Dalal, N.; Trigggs, B. Histograms of oriented gradient for human detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, San Diego, CA, USA, 20–26 June 2005; pp. 886–893.