



An Interpretability Case Study of Unknown Unknowns Taking Clothes Image Classification CNNs as an Example

Huan Li and Yue Wang^(✉)

School of Information, Central University of Finance and Economics, Beijing, China
yuelwang@163.com

Abstract. “Unknown unknowns” are instances predicted models assign incorrect labels with high confidence, greatly reducing the generalization ability of models. In practical applications, unknown unknowns may lead to significant decision-making mistakes and reduce the application value of models. As unknown unknowns are agnostic to models, it is extremely difficult to figure out why models would make highly confident but incorrect predictions. In this paper, based on identification of unknown unknowns, we investigate the interpretability of unknown unknowns arising from convolutional neural network models in image classification tasks by interpretable methods. We employ visualization methods to interpret prediction results on unknown unknowns, further understand predictive models and analyze the predictive basis of unknown unknowns. We focus the application scenario of interpretability of unknown unknowns on a clothes category recognition task (dress vs shorts) in e-commerce platforms, and observe some patterns of models making wrong classifications that lead to unknown unknowns, which indicates that a CNN model that lacks of common sense can make mistakes even for a large dataset. Besides, we observe some interesting phenomena: certain correct predictions of instances are unreliable due to wrongly identified features by CNNs.

Keywords: Unknown unknowns · CNN Interpretability · CNN Visualization

1 Introduction

Unknown unknowns refer to image data that are misclassified with high confidence in image classification tasks, revealing the models’ inability to detect these errors. There are various reasons for unknown unknowns, such as the limitation of datasets, the emergence of new categories, etc. “Unknown unknowns” problem can be disastrous in some special application scenarios. e.g., in the medical field, exist where the categories or certain features of these cases are absent from previous datasets. As a result, medical predictive models may fail to diagnose or misdiagnose, leading to significant and tragic consequences in disease decision-making. Therefore, addressing the “unknown unknowns” problem is crucial for enhancing the accuracy and generalizability of predictive models in image classification.

Nowadays image classification has become a focus in the field of machine learning. Convolutional Neural Networks (CNNs) [1] has emerged as a classic and high-performing model for image classification. As image classification continues to evolve, model interpretability has garnered widespread attention. Interpretability research aims to convert the output of black-box deep learning models' image predictions into human-understanding formats, using specific methods and techniques. Despite significant progress, the interpretability research faces challenges, one of which is unknown unknowns, resulting from factors like incomplete datasets and poorly extracted features. Due to their negative impact on apparel classification in e-commerce platforms, it is crucial to ensure accurate and reliable image recognition technology. Hence, we focus on clothes image classification in e-commerce platforms as a specific application scenario for our interpretability research.

To tackle unknown unknowns in clothes classification, we train two CNN model (VGG [2] and ResNet [3]), identify the unknown unknowns and visualize their prediction results using interpretable methods, which allows us to gain insights into their inner mechanisms. We also compare the performance of two different CNNs. Finally, we investigate how the resolution of image data affects model performance and the occurrence of unknown unknowns.

Using two interpretation methods (Class Activation Mapping [4] and Local Interpretable Model-agnostic Explanations [5]), we uncover valuable facts: unreliable correct predictions of instances (not uncommon for key areas to differ between methods on the same instances), which have been previously overlooked. e.g., an instance is correctly classified as "shorts", but the CAM result highlights the vest as the key area instead of the shorts. This unreliable correct prediction is probably attributed to similar edge characteristics between them. Additionally, models may erroneously focus on irrelevant aspects, such as distinct edges or human body parts, as key areas.

2 Related Work

This section provides an overview of prior research on CNN semantic problems in identifying unknown unknowns. The distinction of previous work is that we identify the semantics problems of interpretability areas for unknown unknowns (images correctly predicted but having unknown wrongly classification features).

2.1 Semantics Problems in CNNs

Network dissection is a pioneer paper which investigates the roles of neurons in CNNs [6]. Following this line, Fong et al. [7] find that in most cases multiple neurons encode a concept, and a single neuron can encode multiple concepts. Mu et al. [8] employ beam search to generate logical forms of primitive concepts and investigate their connection to neurons. They discover that in image classification, some neurons learn highly abstract and semantically coherent visual concepts, while others detect unrelated features. Olah et al. [9] propose a microscopic approach to studying interpretability by carefully examining neurons and circuits, similar to using microscopes to study microorganisms in

history. Surprisingly, they report an instance where a car detector spreads its car feature to a dog detector in the next layer. They also observe equivariance, a term borrowed from biology, where multiple neurons can detect different posed dog faces [10]. Hohman et al. [11] introduce an interactive system called SUMMIT that provides a summary and visualization of the features learned by a deep learning model.

2.2 Identify Unknown Unknowns

Lakkaraju et al. [12] propose a two-phase method using descriptive space partitioning (DSP) and Bandit for Unknown Unknowns (UUB), which shows progress in semi-automatic identification. Bansal et al. [13] introduce a utility model based on coverage, encouraging exploration in high-density regions not adjacent to discovered unknown unknowns. Compared to DSP + UUP, this method discovers diverse unknown unknowns and achieves a more evenly distributed effect in their discovery. Subsequently, Dong and Dong et al. [14] present a region selection model using unsupervised learning for improved generalization and robustness in image classification tasks.

2.3 Improve Accuracy for Specific Application Scenarios

Various novel methods have been employed to improve classification accuracy in specific image classification applications. The FSCAP model [15] consists of multiple functional modules to enhance the accuracy of fashion sub-category and attribute prediction. Shajini et al. [16] propose an attention-driven technique that enables the model to capture multiscale contextual information of landmarks, thereby improving classification performance. Li et al. [17] utilize a teacher–student (T–S) pair model in a semi-supervised multi-task learning approach on unlabeled clothing datasets. Additionally, a multimodel cascaded convolutional neural network (MCCNN) [18] is introduced for garbage classification, effectively suppressing false-positive predicts.

3 Experiments

We now present the details of the experiments of training model, discovering unknown unknowns and visualizing the interpretation results.

3.1 Experiment Preparation

In our experiment, we use the DeepFashion dataset [19], which offers detailed classification with numerous categories. e.g., the original dataset includes specific dress categories based on features like tightness, prints, plaid, logos, and dress styles. However, the original dataset contains a limited number of images per category (around 20 to 50), which is insufficient for clothes recognition requiring a larger number of samples.

To adapt the DeepFashion dataset for our application, we integrate and reorganize the image data. We consolidate subcategories within the same clothes category, ensure correct labels, and renumber the images based on their sequence, thereby creating a binary classification dataset comprised of approximately 5000 image samples, with a focus on the categories of dresses and shorts.

3.2 Experiment Process

Construct Two CNNs and Evaluate Their Performance. We construct and train VGG16 [2] and ResNet18 [3], both known for their excellent performance. Additionally, we add an adaptive average pooling layer (AdaptiveAvgPool2d) to VGG16 for subsequent visualization research using the CAM method.

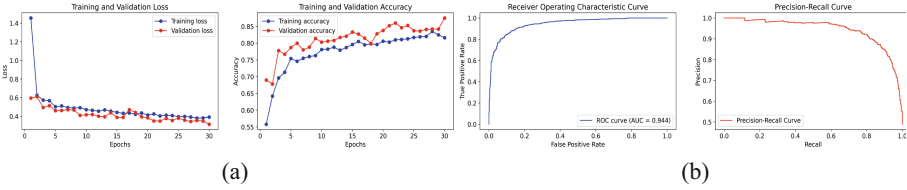


Fig. 1. (a): Loss and accuracy of VGG16 on training set and test set; (b): ROC curve and PR curve of VGG16 on test set.

As shown in Fig. 1, the model achieves approximately 85% accuracy on the validation set after several iterations. The loss consistently decreases, and the accuracy converges to around 85% from the third round. These results indicate the VGG16 model demonstrates robust classification performance on this dataset.

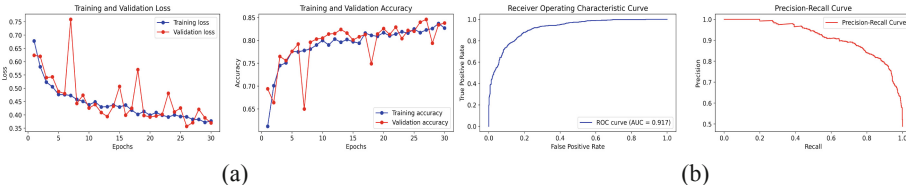


Fig. 2. (a): Loss and accuracy of ResNet18 on training set and test set; (b): ROC curve and PR curve of ResNet18 on test set.

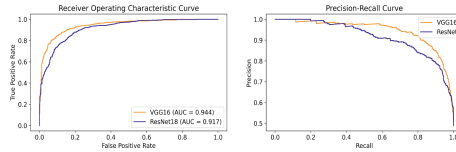
The ResNet18 model achieves the highest accuracy of approximately 84% on the validation set in the final iteration, which is similar to VGG16. Figure 2 shows significant fluctuations in loss and accuracy on both sets during the 7th and 18th iterations, but overall, the model’s classification performance steadily improves.

Compare Performance of Two Models in Clothes Classification. Table 1 shows VGG16 outperforms ResNet18 in AUC, Accuracy, F1 score, and Precision, though ResNet18 has a slightly better Recall score. Figure 3 indicates a larger area under the ROC and PR curves for VGG16 compared to ResNet18, demonstrating that VGG16 exhibits better classification performance on this dataset. Consequently, we use the prediction results of VGG16 for subsequent interpretability research.

According to the reference paper, unknown unknowns are misclassified image data with high confidence. For our experiment, we set a confidence threshold (α) of 0.65 to identify unknown unknowns.

Table 1. Various classification performance indicators of VGG16 and ResNet18.

Indicators	ResNet18	VGG16
AUC	0.917	0.944
Accuracy	0.794	0.812
Precision	0.788	0.820
Recall	0.811	0.794
F1	0.787	0.802

**Fig. 3.** ROC curve and PR curve of VGG16 and ResNet18.

Employ two Interpretability Methods for Prediction Visualization. Based on the pretrained VGG16, we implement the CAM method [4] to obtain a weighted average feature map where each pixel represents the intensity of the effect of that location on the target category. A heat map is created by retaining the pixels with values over zero, normalizing them, and scaling to the original image size. Finally, the heat map is converted to RGB format and superimposed onto the original image to visualize significantly influential areas.

Additionally, we employ the LIME method [5] for visual interpretation. We obtain the output vector and prediction result by inputting image data into the pretrained model. We construct the LIME image interpreter, adjust parameters, and apply the resulting mask to the original image for visualization. LIME is a model-agnostic method that can be applied to any CNN model, such as VGG16. However, it necessitates the definition of a prediction function that is compatible with LIME.

4 Results

4.1 Interpretable Results of CAM Method

Visualize and Analyze the Visualization Results of Two Unknown Unknowns. The visualization result is obtained by overlaying the CAM diagram and the original image. (a) displays the original heat map. (b) displays the heat map converted to RGB. (c) displays the original image. And (d) displays the overlay diagram of the heat map and the original image. Different colors in (d) represent varying levels of attention given by the model to different areas of the image during classification. Colors closer to red indicate areas where the model pays more attention.

The true label of unknown unknown A is “dress”, but the model predicts it as “shorts” with a confidence level of up to 82%. In Fig. 4(A)–(d), the model focuses more on the

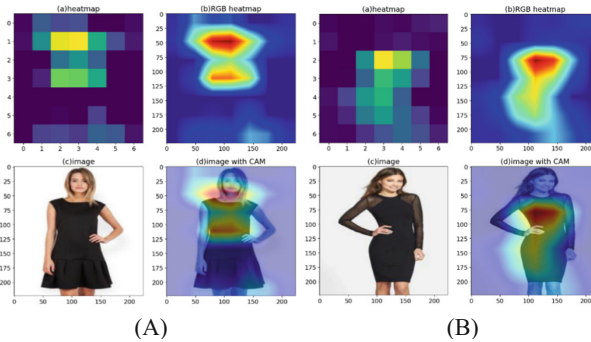


Fig. 4. (A): Visualization result of unknown unknown A with CAM; (B): Visualization result of unknown unknown B with CAM. For the figures of (A) or (B): (a) top left, (b) top right, (c) bottom left, (d) bottom right.

shoulders and abdomen of the portrait, while bright-colored areas are mainly in the upper part of the clothes. This indicates A’s key area interpreted by CAM is concentrated in the upper part of the clothes, whereas the lower part of the clothes is actually crucial for recognizing whether it is a dress. The misalignment between the key area and the actual distinguishing features of clothes could be latent reason for the occurrence of unknown unknowns.

















The true label of unknown unknown B is “dress”, but the model incorrectly predicts it as “shorts” with a confidence level of 93%. In Fig. 4(B)–(d), the red area represents the chest of the portrait, while the bright-colored area is mainly distributed in the upper part and right edge of the dress. The model focuses on most areas of the clothes for the prediction but gives more attention to the upper part, similar to unknown unknown A. The model also considers the edge of the lower part of the clothes. However, due to the tight-fitting nature of the dress, it may exhibit characteristics resembling the edge lines of shorts, which could be another possible reason for unknown unknowns.

Sample and Compare Visualized Results of Unknown Unknowns with the Correctly Classified Data by CAM Method. 11 unknown unknowns and 5 correctly classified data are randomly sampled for visual comparison, which makes the experiment more rigorous and comprehensive.

In Table 2, the key areas of unknown unknowns 1 and 2, similar to A and B, concentrate on the upper part of clothes. Unknown unknown 3’s key area is limited to the hem of the clothes, suggesting a poor capture of overall features. The key area of unknown unknown 4 is the environment with prominent edges, unrelated to clothes features. The key areas of unknown unknowns 5, 6, 7, 8, and 9 are clothes edges close to the environment, indicating sensitivity to obvious edge features. Unknown unknown 10 focuses on the hem of clothes and human body, which resembles the unknown unknown 3.

For correctly classified data, key area of data 1 remains at the dress edge, correctly predicting it as “dress”. Unknown unknowns 5, 6, 7, 8, and 9 also focus on the dress edge but incorrectly predict shorts as “dress”, which is interesting finding that a potential relationship between the dress edge and the model’s predictions of “dress”. The key area of data 2 covers most of the clothes, obtaining the correct prediction. Though

Table 2. Visual results of CAM method for unknown unknowns and classified correct data sampling (UU for unknown unknowns, CC for correctly classified data).

Type	N	Interpretation results	The key areas	Type	N	Interpretation results	The key areas
	O.				O.		
UU	1		Upper part of clothes	UU	9		Edge of clothes and human body
UU	2		Upper part of clothes	UU	10		Hem and human body
UU	3		Hem	UU	11		Upper part of clothes
UU	4		Environment	CC	1		Edge of clothes
UU	5		Environment and edge of clothes	CC	2		Edge of clothes
UU	6		Edge of clothes and human body	CC	3		Vest
UU	7		Edge of clothes and human body	CC	4		Shorts
UU	8		Edge of clothes and environment	CC	5		Shorts

correctly classified, the model predicts data 3 as “shorts”, with its key area being the vest, suggesting some unreliable prediction results for correctly classified data. The key areas of data 4 and 5 represent shorts, consistent with their prediction results. These two predictions are reliable.

4.2 Interpretable Results of LIME Method

Visualize and Analyze the Visualization Results of Two Unknown Unknowns. We use LIME method to visualize model’s prediction results. (a) displays the top 2 features with significant impact. (b) displays the top 5 features with significant impact. (c)

displays the top 5 features comprehensively considering positive and negative contributions. Positive features contribute to results (e.g., the clothes parts), while negative features contribute nothing or may confound (e.g., the human body parts). (d) displays a visualization where features with weights < 0.1 are excluded, thereby omitting irrelevant features.

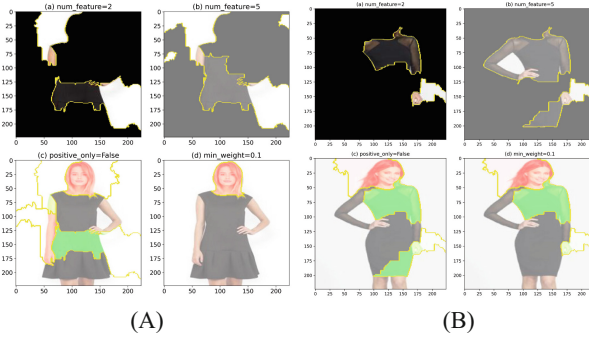



Fig. 5. (A): Visualization result of unknown unknown A with LIME; (B): Visualization result of unknown unknown B with LIME. For the figures of (A) or (B): (a) top left, (b) top right, (c) bottom left, (d) bottom right.

Figure 5(A)-(a) shows that the key areas are the middle part of the dress and the environment. Whereas key area in Fig. 5(A)-(b) covers larger clothes area than that of Fig. 5(A)-(a), indicating poor performance in extracting the first two features. In Fig. 5(A)-(c), positive and negative features are illustrated, where the green area represents the positive contribution to the prediction result, mainly focusing on the middle part of the dress, while the red area covers the face, which is not a significant factor in differentiating between dresses and shorts. When examining Fig. 5(A)-(d), only the human face, which represents the negative area, has feature with weight > 0.1 . It should be noted that the interpretable results of the LIME method for A are not entirely consistent with those obtained from the CAM method.

In Fig. 5(B)-(a), the top 2 important features, the upper part of the clothes and the human hand, are similar to those interpreted by the CAM method for B. Compared to Fig. 5(B)-(a), Fig. 5(B)-(b) extends the key area to include the left edge of the clothes. Figure 5(B)-(c) shows more comprehensive interpretation, where the green area represents interpretable positive features for classification, and the red area represents interpretable negative features, indicating that the area of the human head is irrelevant to the classification result. When we ignore features with weights < 0.1 in Fig. 5(B)-(d), the green area still highlights the model's focus on the upper part of the clothes instead of the lower part.

Sample and Compare Visualized Results of Unknown Unknowns with the Correctly Classified Data by LIME Method. We interpret the prediction results of 16 instances sampled, including 11 unknown unknowns and 5 correctly classified data, using the LIME method. We also analyze and compare the results of LIME and CAM.

Table 3. Visual results of LIME method for unknown unknowns and classified correct data sampling (UU for unknown unknowns, CC for correctly classified data).

Type	N	Interpretation	The key areas	Type	N	Interpretation	The key areas
	O.	results			O.	results	
UU	1		Left side of clothes and environment	UU	9		Edge of clothes and environment
UU	2		Waist	UU	10		Left side of clothes
UU	3		Right side of clothes	UU	11		Part of clothes
UU	4		Lower part of clothes	CC	1		Human body
UU	5		Middle part of the clothes	CC	2		Left side of clothes
UU	6		Upper part of clothes	CC	3		Vest and environment
UU	7		Left side of clothes and environment	CC	4		Environment and vest
UU	8		Part of clothes and environment	CC	5		Shorts and human body

In Table 3, the key area for unknown unknown 1 is the left side of the dress and the environment. However, CAM result for unknown unknowns 1 focus on the upper part of the dress, which is different from LIME result. The key area for unknown unknown 2 is the waist of the dress, while CAM result is the upper part of dress, indicating both methods interpret similar key areas. The key area for unknown unknown 3 is only the right side of the dress. But CAM result is the hem of the dress. There is no overlap between the key areas interpreted by the two methods. The key area for unknown unknown 4 is the lower part of the dress. It significantly differs from the CAM result (the environment). The key area for unknown unknown 5 is the body of the dress, covers most but not edges. Despite the contribution of result to the prediction decision, unknown unknown 5 is incorrectly

predicted. The key area for unknown unknown 6 is the upper part of the clothes, while the key area of CAM is the edge of the clothes and the human body. Both sets of results indicate that the model focuses on the upper part of the clothes when classifying. The key area for unknown unknown 7 is the left side of the shorts and the environment. Yet, CAM result for unknown unknown 7 is the edge of the clothes and the human body. The overlapping key areas are located at the left edge of the vest, confirming that the model focuses on edge features. The key areas for unknown unknown 8 is the body of the clothes and the environment, including part of the edge, which is the same as the key areas interpreted by the CAM. The key area for unknown unknown 9 is the environment close to dress edge, which matches CAM method result, confirming the model focuses on edge again. The key area for unknown unknown 10 is the right part of the dress, but CAM result is the hem of the dress and the human body. The key area for unknown unknown 11 is a small portion of middle of dress, differs from CAM method's key area as upper part of dress.

The key area for correctly classified data 1 is the human body, whereas the CAM result focuses on the edge of the clothes. They are completely different results. We suspect that the top 2 features are insufficient for a comprehensive interpretation or that the model is unable to extract key features effectively. The key area for data 2 is the left side of the clothes, differs from CAM result that covers most of the clothes. The key area for data 3 is the upper part of the clothes and the environment, which is consistent with CAM result. However, the expected key area is the lower part of the clothes, indicating that the prediction results may be unreliable. The key area for data 4 is the human body and the upper part of the clothes, different from CAM method. Similar to correctly classified data 1, LIME result on correctly classified data 4 is unrelated to shorts but predicted correctly. There are possible reasons: unreliable LIME interpretations (different from CAM results) or inappropriate selection of the top 2 features (more features needed for interpretation). The key area data 5 is the shorts and the human body, same as CAM result. The model seems to focus excessively on shape of the human body, as the key area.

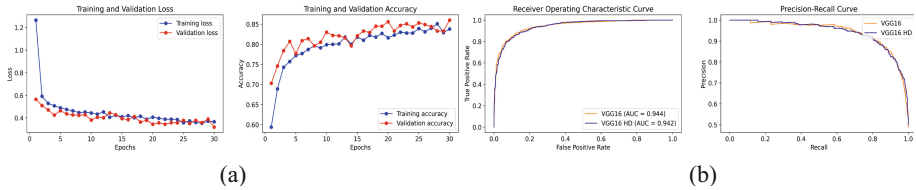
4.3 Impact of Image Resolution on Classification Performance and Unknown Unknowns

In the initial experiments, we use low-resolution images. However, high-resolution images provide more information and yield better classification performance during model training. To investigate the impact of image resolution on classification performance and unknown unknowns, experiments are performed using HD image data from the same dataset, processed in a similar manner.

As shown in Fig. 6, the model achieves an accuracy of 86% on the high-resolution dataset, slightly higher than that on the low-resolution dataset by one percentage point after around 30 rounds of convergence. Table 4 shows the VGG16 model performs better in terms of AUC and recall scores, while the VGG16 HD model exhibits higher accuracy, F1 score, and precision score. The ROC and PR curves in Fig. 6 demonstrate similarities between the two models. Overall, image resolution appears to have minimal impact on the models' classification performance in this particular task. Additionally,

Table 4. Various classification performance indicators of VGG16 on low-resolution image data and VGG16 HD on high-resolution image data.

Indicators	VGG16 HD	VGG16
AUC	0.942	0.944
Accuracy	0.820	0.812
Precision	0.815	0.820
Recall	0.721	0.794
F1	0.818	0.802

**Fig. 6.** (a): Loss and accuracy of VGG16 HD model on high-resolution training set and test set; (b): ROC curve and PR curve of VGG16 and VGG16 HD on HD image data.

when sampling unknown unknowns from the VGG HD model, we find overlap with those from the VGG model, suggesting no significant correlation between image resolution and unknown unknowns. However, further rigorous experiments are necessary to establish conclusive results.

5 Conclusion and Future Work

In our study, we find some interesting facts that we believe are valuable to the community of CNN researchers taking the clothes classification CNNs in E-Commerce as an example: We explain unknown unknowns by classic interpretability methods CAM and LIME, that is, the correct predictions of some instances are not always reliable. e.g., a pair of shorts is predicted correctly, but the key part interpreted by CAM or LIME is that of a vest. This confusion may be due to similar edges features present in both items, as evidenced by some visualized image samples. Our work contributes to a deeper understanding of the internal mechanisms behind the predictive decisions made by black-box CNN models, showing their lack of common sense from the perspective of unknown unknowns. We test VGG and ResNet CNNs which are widely used and trusted in the current practice (especially ResNet) due to good classification performances. However, the interpretive results highlight issues with these models when they are used solely for binary classification tasks (dress vs shorts). And we use the standard interpretation models CAM and LIME. They are widely used in current research and have been demonstrated to deliver reasonably good interpretative performance across various studies. In the future, we shall test more recent CNN models such SOTA by some new designed interpretation models.

Acknowledgements. This work is supported by: National Defense Science and Technology Innovation Special Zone Project (No. 18-163-11-ZT-002-045-04); Engineering Research Center of State Financial Security, Ministry of Education, Central University of Finance and Economics, Beijing, 102206, China. The code is at: https://github.com/marcherwang/cgi2023_unknown_unknown_paper.

References

1. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proc. IEEE* **86**(11), 2278–2324 (1998)
2. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014)
3. 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* (2016)
4. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016)
5. Ribeiro, M.T., Singh, S., Guestrin, C.: “Why should I trust you?” Explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016)
6. Bau, D., Zhou, B., Khosla, A., Oliva, A., Torralba, A.: Network dissection: quantifying interpretability of deep visual representations. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017)
7. Fong, R., Vedaldi, A.: Net2vec: quantifying and explaining how concepts are encoded by filters in deep neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018)
8. Mu, J., Andreas, J.: Compositional explanations of neurons. *Adv. Neural. Inf. Process. Syst.* **33**, 17153–17163 (2020)
9. Olah, C., Cammarata, N., Schubert, L., Goh, G., Petrov, M., Carter, S.: Zoom in: An introduction to circuits. *Distill* **5**(3), e00024-001 (2020)
10. Olah, C., Cammarata, N., Voss, C., Schubert, L., Goh, G.: Naturally occurring equivariance in neural networks. *Distill* **5**(12), e00024-004 (2020)
11. Hohman, F., Park, H., Robinson, C., Chau, D.H.P.: Summit: scaling deep learning interpretability by visualizing activation and attribution summarizations. *IEEE Trans. Visual Comput. Graphics* **26**(1), 1096–1106 (2019)
12. Lakkaraju, H., Kamar, E., Caruana, R., Horvitz, E.: Identifying unknown unknowns in the open world: Representations and policies for guided exploration. *Proc. AAAI Conf. Artif. Intell.* **31**(1), 2125–2132 (2017)
13. Bansal, G., Weld, D.: A coverage-based utility model for identifying unknown unknowns. *Proc. AAAI Conf. Artif. Intell.* **32**(1) (2018)
14. Dong, X., Zhang, H., Demartini, G.: A region selection model to identify unknown unknowns in image datasets. In: *ECAI 2020*, pp. 474–481. IOS Press (2020)
15. Amin, M.S., Wang, C., Jabeen, S.: Fashion sub-categories and attributes prediction model using deep learning. *Vis. Comput.* **39**(6), 3851–3864 (2023)
16. Shajini, M., Ramanan, A.: An improved landmark-driven and spatial-channel attentive convolutional neural network for fashion clothes classification. *Vis. Comput.* **37**(6), 1517–1526 (2021)
17. Li, J., et al.: Automatic detection and classification system of domestic waste via multimodel cascaded convolutional neural network. *IEEE Trans. Ind. Informatics* **18**(1), 163–173 (2022)

18. Shajini, M., Ramanan, A.: A knowledge-sharing semi-supervised approach for fashion clothes classification and attribute prediction. *Vis. Comput.* **38**, 3551–3561 (2022)
19. Liu, Z., Luo, P., Qiu, S., Wang, X., Tang, X.: Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016)