Zero-Shot Learning-Based Detection of Electric Insulators in the Wild

Ibraheem Azeem and Moayid Ali Zaidi

May 4, 2021

Zero-Shot Learning-Based Detection of Electric Insulators in The Wild

Ibraheem Azeem*, Moiyad Ali Zaidi*
Østfold University College Halden, Norway

Abstract

An electric insulator is an essential device for an electric power system. Therefore, maintenance of insulators on electric poles has vital importance. Unmanned Aerial Vehicle (UAV) are used to inspect conditions of electric insulators placed in remote and hostile terrains where human inspection is not possible. Insulators vary in terms of physical appearance and hence the insulator detection technology present in the UAV in principle should be able to identify a insulator device in the wild, even though it has never seen that particular type of insulator before. To address this problem a Zero-Shot Learning-based technique is proposed that can detect an insulator device type that it has never seen during the training phase. Different convolutional neural network models are used for feature extraction and are coupled with various signature attributes to detect an unseen insulator type. Experimental results show that inceptionsV3 has better performance on electric insulators dataset and basic signature attributes "Color and number of plates" of the insulator is the best way to classify insulators dataset while the number of training classes doesn't have much effect on performance. Encouraging results were obtained.

Keywords: Zero-shot learning, signature attribute, electrical insulators, object detection

1 Introduction

Deep learning has shown incredible performance in problms like image classification and object detection, yet this comes at the cost of a huge number of annotated samples for training. It is also not possible to correctly identify the unseen classes those are not used while training and hence in such situation we need to train the classifier from scratch again. In the past few years, special forms of neural networks "Convolutional Neural Networks (CNNs)" are advancing the state of the art of the many computer vision applications, and these are specifically designed to require advantage of the 2D structure of image data. However, CNNs typically require an enormous amount of training data as well that is usually not available in many tasks such as power cables component detection and classification [10]. To the best of our knowledge, the very little approach is available that might efficaciously ankle these quandaries across many tasks. Some of the approaches to handle the dearth of training data problems are i. Utilizing synthetic data and data augmentation techniques to come up with more training data ii. Employing one-shot learning that aims to be told information about incipient object categories from just one or some training images [12] iii. Adopting zero-shot learning that seeks to ascertain information about incipient object categories from only descriptions [15].

Zero-shot learning is used to generalize the learned information to explore the unfamiliar object classes. During this context, both unseen and seen classes require to be cognate to the auxiliary semantic features. The setting within zero-shot learning considered the profound case of transfer learning. The model is trained to replicate human competency on unseen classes, which don't seem to be within the training stage [4]. The core in zero-shot learning is to ascertain a multi-model projection between semantics and visual features by utilizing the labeled optically discerned classes [8]. Besides this, the patrol transmission line of Unmanned Aerial Vehicle (UAV) has become very popular in transmission line inspection and research hot topic. It has shown efficiency, reliability, cost-effectiveness, and helped in inspections where the human approach is not easy. With the development of automation in UAV, it's the interest to identify the types of insulators during an inspection by capturing images of the insulator by UAV camera. For this purpose, a large number of images of insulators are required which is always a problem in deep learning. In this paper, we are addressing this problem by introducing a zero-shot model and use insulators dataset images captured by UAV.

Considering the objectives involved in the study, our research comes up with the subsequent questions: 1. Which deep-architecture is able to generate the most powerful features? 2. What type of "Signature Attributes" can we propose to perform ZSL for our defined problem? 3. Does the number of training classes affect system performance?

2 Related Work

Researchers have used different approaches to machine learning to achieve a lack of data training problems. The have proposed multiple ways in their studies to apply zero-shot learning on the datasets.

Study in [8] introduces a multi-model explication model that categorically integrates three LSTMs (long short-term recollection models) and implemented on CUB, SUN, and AWA2 dataset. It shows the way to extract visual and textual explications. Furthermore, two incipient aspects are insights as well i.e. explanatory diversity and explanatory consistency respectively. Accuracy on both optically discerned and unseen classes was 38.4 and 35.6. In [9] zero-shot learning Hierarchical relegation approach for previously unseen classes by trading specificity for accuracy and mapping to semantic attributes of unseen classes from image features to human recognizable attributes are utilized. The direct method for semantic attributes discussed in [2] shows that the classifier is learned for every attribute from the examples within the training dataset. The study in [16] used binary embedding predicated zero-shot learning (BZSL) that apperceived an unseen instance by efficiently probing its most proximate class codes with minimum Hamming distance while incrementing the binary embedding dimension can inevitably improve the apperception precision. [5] has proposed the hybrid model consists of "Random Attribute Selection" (RAS) and conditional "Generative Adversarial Network" (GCN) to find out the realistic generation of attributes by their correlations in nature and improve discrimination for a large number of classes while [14] showed an approach for learning semantic-driven attributes using two different datasets i.e. AwA and aPY. Human gaze embedding is used as auxiliary information to be told compatibility between image and label space for zero-shot learning [6]. Embedding framework that maps multiple text parts is joined with multiple semantic parts into a typical space in [1]. Moreover, Fine-grained Caltech UCSD Birds-2011 dataset is used and improves state-of-the-art on the CUB dataset to 56.5 (from 50.2) within the supervised setting while improving the state-of-the-art also within the unsupervised setting to 33.9 (from 24.2). In [13] AwA dataset is used as in [14] also, contains 30,475 images from 50 animal classes while each class is annotated with 85 user-defined attributes. CUB dataset contains 11,788 images of 200 bird classes and each class is annotated with 312 attributes. User-defined semantic attributes jointly with discriminative and background lat1ent attributes approach are proposed in the study. The limitations of the proposed approach are that it takes a fixed feature representation as an input. A joint feature and attribute learning approach is suggested to overcome the limitations on some level. Another study [11] has used a semantic attention-based comparative network (SACN) to resolve the zero-shot visual recognition problem with the same dataset state above. ResNet and GoogleNet models are used as a backbone in it. The best results reported in the study are 86.5 on AWA using ResNet as a backbone while GoogleNet gives 84.5. For the CUB dataset 63.4 and 62.0 are reported results respectively. [18] has used both Nearest Neighbor + Self-Training (NN + ST) and data augmentation on video action recognition tasks. However, simple self-training and data augmentation strategies can address these challenges and achieve the state of the art results for zero-shot action recognition within the video. The study showed that Semantic embedding is similar to the state of the art low-level feature-based classification and better than the standard attribute-based intermediate representation. The possible reasons given are, attribute-space being less discriminative than semantic word space used, or because of the reliance on human annotation. This shows that some annotated signature attributes are might not be detectable or don't seem to be discriminative for class [18]. Objects are identified supported a high-level description that's phrased in terms of semantic attributes, like the object's color or shape. NN isn't better than the attribute-based approaches in [7]. The zero-shot classification in [17] is used to classify target classes specifically supported

 

Figure 1 Figure 2

learning a semantic mapping. It targets classes from feature space to semantic knowledge space. Considering the model as best, it can give the best accuracy in both the AwA dataset and CUB datasets. In the study [3], the researcher has proposed a unified semi-supervised learning framework. This framework learns the attribute classifiers by exploring the correlations between images and utilizing multiple signature attributes. The usage of multiple features makes the attribute prediction more vigorous. An optimal graph is another choice to enhance the execution of zero-shot image categorization whether or not labeled images are less.

A literature study has shown that different types of approaches are proposed to implement zero-shot learning on an identical style of dataset. Most of the researches have used the datasets relevant to animals or birds. No study has used the dataset we have proposed for zero-shot learning. Due to the unusual attribute nature of our dataset, we are required to put it for test and determine the most effective possibilities in it.

3 Dataset Details

The dataset contains different images of insulators that are captured against several types of complex background by UAV camera. The background of image scenes are with the vast majority of different other objects that includes forest, crops, grass, poles, and wires etc. Using UAV camera insulator in images was placed on a special frame so that the images can be obtained in conditions close to the real environment. The images of insulators are structured in folders in such a way that there basic attributes should reflect the difference between them. Figure 1 and Figure 2 are example images of two different insulators.

4 Methodology

In the context of recognizing insulators, one must provide large number of data which is not really available when one uses UAV for power inspection images. Moreover, a deep learning algorithm can classify a test data to any of its training classes, but fail completely when dealing with an “unseen” class. On the other hand, it cannot be ignored that deep learning features extraction are discriminatory in nature which is not acceptable in some cases. Here, the proposed method uses deep learned features along with a "Zero-shot Learning" framework to counter the problem of recognizing "untrained" class of images.

A: Zero-shot Learning

Starting from pattern recognition and the theory that computers can learn without instructions, researchers showed interest to see if computers could learn from data. To classify the samples where training examples are not available, we use zero-shot learning. Recognizing an object from an image without training the image using zero-shot learning allows us to recognize the unseen class of the object. It can give a high-level description of a new or unseen class and create a connection between it. Inspired by this human's ability, the interest of researchers in zero-shot learning is increasing for scaling up visual recognition.

B: Semantic Attributes

Recent advancement in object detection are directly learning a mapping from an image feature space to a semantic space. For that, the semantic attribute approach is the best way to go with it. Semantic attributes provide a bridge from automatically generated image features to human intuition . In our case, we have a dataset of electric insulators and there are several types of insulators. The most commonly used are the pin type, strain insulator, shackle insulator, and suspension type. We aim to differentiate the insulator on the bases of their characteristics and signature attributes. To get information about the attribute we must focus on the naturally shows up region instead of the entire spatial domain. Different objects can have certain similar attributes but their category label is preserved. Attributes are also especially useful in problems that aim at modeling intra-category variations such as fine-grained classification. Even objects are encoded according to their semantic attributes or features and have become quite very practical now a day. An object can be encoded over a large set of exploratory attributes, while each attribute can be assigned a specific or multiple value/s that shows its probability, weight, or importance.

5 Experimental Results

The dataset consists of 38 total number of classes that we distributed to training and validation. We randomly took 30 classes from it to fine-tune the pertained models. The object of the experiments on this stage is to get to know which deep learning architecture can generate the most features from the dataset.

"Which deep-architecture is able to generate the most powerful features?"

We already knew that we have a large number of classes in our dataset but images in each dataset are so less than models will not able to train them in a good manner. We decided to go for testing on the dataset and check, what we get in results. We tried to tune VGG16, VGG19, Xception, and inceptionV3 models on the dataset and got expected results as we assumed before. We did not get results better than 0.10, which we got from VGG16. We decided to go for data augmentation on this stag to increase the images in the dataset and tried to deceive the models. There are multiple approaches proposed by people for data augmentation. The first approach we used was changing the shape of the images. We increase the number of images in each class from 100 to 600 only one class contained 400 images because its original images were so less i.e. 50. Original images size in the dataset was mostly 1397X1397 but some were different as well. After augmentation sizes, all dataset images size, become 240X240 each.

Experiment#1:

We started tuning of models again on random 30 classes and got some different results from the previous experiment. We divided the dataset to training and validation with 0.80 for training and 0.20 for validation. Total number epoch was set as 100 with batch size 32 and the input image size is 240 X 240 X 3. We aimed to get the exact bounding box area/cropped of electric insulators so in the "ImageDataGenerator" function we used "shear_range" and "zoom_range" as 0.1. We also did "horizontal_flip" as True on the images. As a result, vgg16 gave 0.43 validation accuracy with 1.56 validation loss with a default learning rate of "Adam" as an optimizer. On decreasing the learning rate of vgg16 validation accuracy increases very slowly for-example it changes in points 0.21, 0.219 then 0.22 so this is the learning progress of vgg16.

Experiment#2:

VGG16 achieved 0.52 accuracy on validation on changing the "shear_range" and "zoom_range" as 0.2. Inceptionv3 did not give any closer result to vgg16. It trained at all on 0.1, 0.01, 0.001 and 0.0001 learning rates. Vgg19 does not give any good results if we used sharp and zoom on the training set but if we removed it, it started learning and gave 0.52 accuracy at the end.

Experiment#3:

We have generated a new dataset using multiple augmentation approaches.

1. Translation
2. Rotation
3. Transformation

In these approaches, we have changed the values of each image degree, rows, and columns. After that, we have refined the dataset again manually and remove the black images containing a small description of the image generated from data augmentation. We split the data to the train and validation folder where mostly original data images were in the validation folder and augmented data were in the train folder. The first observation on tuning did not improve the results we got before but later we mix the randomly in training and validation folders again and then run the experiment. We also changed some of the parameters for the training model. Set the learning rates (0.0001, 0.001), beta_1=0.9, beta_2=0.999, epsilon=1e-06 and amsgrad=True of Adam optimizer and set all zoom properties to 0.2. The results we got on different models are given in Table 1.

CNNs	val_accuracy
InceptionV3	75.5%
Xception	72.4%
VGG16	68.7%
VGG16	69.4%

Table 1: CNN Results

No	Attributes
1	Color, Selected body part, No. of plates
2	Color, Selected body part No. of plates, Plates linked on body
3	Color, Selected body part, No. of plates Plates linked on body, Other Selected part, Shape of other selected part
4	Color, Selected body part, No. of plates Plates linked on body Other Selected part, Shape of other selected part, Surface
5	Color, Selected body part, No. of plates Selected part1, Shape of part1, Selected part2, Shape of part2, Surface

Table 2: Signature Attributes

"What type of "Signature Attributes" can we propose to perform ZSL for our defined problem?"

To classify the data for zero-shot learning, we arrange the data into their homogeneous/similar groups according to the common characteristics they have. Data without a proper classification is not understandable by CNN and it is not valuable for further analysis and interpretation. The arrangement of data will helps CNN in training and provide ease to do comparison and analysis. We have different types of insulators in our dataset that can be defined concerning their properties, feature, and physical appearance, see Table 2.

Observations: We can create more combinations of attributes and define the objects with more details for classification. Some attributes are common in most of the classes, for-example our dataset contains one class of insulator that has brown color but there are certainly other classes as well that has the same property. We need can separate the attributes on the bases of colors or the combination of other attributes as well. In general, our dataset is classified on the bases of two attributes that are color and number of plates that give us 38 classes in total. On selection of certain attributes and going into more detail of objects we can either reduce or increase the number of classes, this was the zero-shot learning come. We are selecting one of the approaches where our data is not based on the number of plates but there are certainly other attributes that we have selected e.g. we have selected colors and how plates are linked on the body.

We have a total number of 11 classes that are created on the bases of brown color and number of plates at the start but now we have selected the above given two signature attributes and the number of classes is reduced to two. In the same way, the total number of 38 classes is now 8 after the new classification of image objects. Now we have classified in such a way that we can use multiple signature attributes and distribution of dataset can show more details. We have chosen color brown, selected body part, number of plates, and selected part with shape and generated new classification by naming the classes like this:

1. Brown_round_top_seven_plate
2. white_long_top_four_plate
3. white_round_neck_eight_plate
4. brown_round_top_six_plate

Using this technique, we have generated 38 classes of the dataset, defining more details of each object is separating it from other objects for classification.

Preprocessing: Here we have used class vector and image embedding technique (CVIE) to check the zero-shot learning score. We generated vectors of both train and zero-shot classes. We considered the word2vec Google news model to generate class vectors, which generated the vector of each class with 300 dimensions. For image embedding with classes, we have used the VGG16 pre-trained model for feature extraction. After that, we have created two files i.e. class vector and zero-shot data files. Class vector file contained an array of all classes whereas zero-shot data contained classes and subclass labels as signature attributes attached with images array. One file in txt format contained train classes that checked when the dataset was training on the bases of their signature attributes. The dataset was partitioned into three i.e. training, validation, and testing. Training and validation contain the data of training classes while testing contained zero-shot classes that were not seen by the model before and it was used to evaluate the zero-shot score on the trained model.

Dataset Training setup: We have used two different activation functions in training i.e. relu and softmax. We have used batch normalization that allows each layer of a NN to learn by itself independently. We have used a "categorical cross-entropy" loss to train the CNNs to output the probability over the classes for each image.

Evaluation of the model: For evaluation of the model, we have used KDTree to query the zero-shot classes on trained classes in our trained model. We have categorized the results into three i.e. top 5, top 3, and top 1 accuracy of the zero-shot model.

Experiment#1: We have used on a large number of attributes for the classification i.e. Basic signature attributes with top details and surface attribute. E.g. 1. brown_round_top_five_plate_solid, 2. glass_pin_top_two_plate_transparent, 3. orange_round_top_thirty_plate_solid. The total number of classes we got on the bases of the above-given classification is 38. We took 32 classes as seen and 6 for unseen classes. To find out zero-shot learning possible score, we set the value "K = 5" to query the results and the best result we got with BATCH_SIZE

Seen classes	Unseen-classes	Top 5	Top 3	Top 1
25	13	0.40	0.23	0.11
33	5	0.47	0.33	0.23
16	22	0.41	0.26	0.13

Table 3:

= 32 and Epoch = 33 were Top_5 Accuracy: 0.37 Top_3 Accuracy: 0.29 Top_1 Accuracy: 0.13

Experiment#2: We have the classified the data based on attributes i.e. Basic signature attributes with top details and surface attribute. The total number of classes we got on the bases of the above-given classification is 38. We took 32 classes as seen and 6 for unseen classes. E.g. 1. brown_round_top_five_plate, 2. glass_pin_top_two_plate, 3. orange_round_top_thirty_plate. To find out zero-shot learning possible score we set the value "K = 5" to query the results and the best result we got with BATCH_SIZE = 32 and Epoch = 53 were: Top_5 Accuracy: 0.38 Top_3 Accuracy: 0.31 Top_1 Accuracy: 0.14

Experiment#3: We have classified the data based on attributes i.e. Basic signature attributes E.g. 1. brown_five_plate, 2. glass_two_plate, 3. orange_thirty_plate. The total number of classes we got on the bases of the above-given classification is 36. We took 26 classes as seen and 10 for unseen classes. To find out zero-shot learning possible score we set the value "K = 5" to query the results and the best result we got with BATCH_SIZE = 32 and Epoch = 53 were: Top_5 Accuracy: 0.54 Top_3 Accuracy: 0.43 Top_1 Accuracy: 0.26

Does the number of training classes affect system performance?

We have done the number of experiments to check the performance of the model by increasing and decreasing the number of classes. In the first phase, we selected the signature attributes that were used in research question#2 i.e. experiment#2 to increase the performance of the model. We did three experiments and chose 23 unseen and 25 seen classes in the first experiment. As a result, we found a slight increase in accuracy for top-5 but a decrease in top-3 and 1. We increase the seen classes to 33 and decreases unseen classes to 5, we observed a good effect on all top-1, 3, and 5 accuracy scores. We reduced the seen classes to 16 and unseen classes to 22 to confirm the effect on performance. Below given table shows the experimental results of the first phase:

In the second phase, we again did three experiments using experimental attributes as used in research question#2 experiment#3, we choose 25 seen and 23 unseen classes. As a result, we found a slight increase in accuracy for top-5 but a decrease in top-3 and 1. We picked 31 seen classes and increases unseen classes to 6 and we observed little incremental effect on all top-1, and 5 accuracy scores but top-3 remained the same. We reduced the classed to seen classes to 16 and used 22 unseen classes to confirm the effect of classes. Below given table shows the experimental results of the first phase:

Seen classes	Unseen-classes	Top 5	Top 3	Top 1
25	13	0.52	0.33	0.13
31	5	0.55	0.43	0.29
16	22	0.47	0.28	0.10

Table 4:

6 Discussion

Different CNN models are used in the study to find out the most powerful feature extraction model on the insulators dataset. On performing three different experiments, we can see that results differ every time because of certain reasons. In the beginning, we can observe that our dataset was so small and with numerous classes, none of the models trained it at all. It demonstrates that deep learning required a large number of data for training. Our first data augmentation technique that only changed the shape of the images and increased the number of images in the dataset was is not helpful to get good training results. On changing parameters like learning rate, batch size, epoch, and layers, training stopped after 0.43. From our second experiment, we can observe that the "ImageDataGenerator" parameters also affect training. Even, changing the value of shear_range and zoom_range hand effects on training and started increasing accuracy i.e. 0.52 on VGG16, but this is not with every model. Our first augmentation technique reflects that there is not much difference in the real and new images. The only resolution of images is changed and CNN models are not able to give us good results on them. To overcome this problem, we can consider multiple data augmentation techniques i.e. rotation, translation, and transformation of the images as a better way to regenerate the data. We can see CNNs works much better on new data generated in our last experiments. With new data generated, without changing any "ImageDataGererator" parameter from the default, learning rate 0.0001 learning rate works properly on all architectures. Furthermore, beta_1=0.9, beta_2=0.999, epsilon=1e-06 and amsgrad=True of Adam optimizer also has effect on training. The use of "spatial categorical_crossentropy" has clear effects on CNNs training as well; "categorical_crossentropy" is used previously. We perceived inceptionv3 as the most powerful feature extraction model on our dataset as compared to VGG16, VGG19, and Xception on the bases of results but still, there was not much difference between Xception and inceptionv3.

For zero-shot learning (ZSL) selection of signature, attributes are one of the important tasks before we do classification. One straightforward and naive strategy for zero-shot to select attributes is to pick such attributes that give larger information. We can analysis that the number of classes also changes on the bases of attributes in the electric insulator dataset. Although, it is hard to select attributes for electric insulators, as some times insulators are cover with some other type of objects that does not have any role to show uniqueness. The basic signature attributes for such type of dataset are enough to classify them and extract a good amount of information. Based on the experimental results, we demonstrate that different attributes have a different level of information and predictability, thus we cannot treat them equally. After the experiments, we can say, on increasing the number of signature attributes on such a dataset decreases the number of classes and we lose the zero-shot accuracy score while a small number of signature attributes we were getting much better results. Insulators have a different type of shapes, but the main difference between them we found out was the color and number of plates for classification. There is always an impact of choosing training and zero-shot classes on results as well. We analyze the performance of the model during testing by changing the number of classes. There is a change in results using a similar number of classes with a different selection of classes for training and zero-shot. The effects of reducing or increasing the number of classes is depending upon the data. In general, there is no guarantee that if we reduce or increase the number of classes with such type of data we can able to increase the classification accuracy. It will improve the performance if we combine similar classes and apply zero-shot on classes that are never seen by the model. Imagine that we are classifying brown color insulators into four classes based on the different numbers of plates. If we train these classes together and apply zero-shot on white color insulators classification, the overall ZSL score will improve.

7 Conclusion

In this study, we have tuned CNNs popular architecture and applied zero-shot learning on UAV captured images of insulators. We have also checked the performance of the model on the bases of number of classes. The experimental results showed inceptionv3 is the powerful CNN model to extract features for us. We also introduced possible signature attributes for this type of dataset to state of the art and applied zero-shot learning. We used class and image embedding (CVIE) approach to test the possible zero-shot score. From results, we can conclude that insulators dataset accuracy is directly dependent upon the classification of insulators on the bases of semantic attributes. Increasing the number of signature attributes will losses the zero-shot accuracy score while a small number of signature attributes better results were achieved on this type of dataset. The best results we got are classification on basic attributes e.g color and number of plates of insulators. The effects of reducing or increasing the number of classes were depending on the data inside each class. We cannot surely say that if we reduce or increases the number of classes we will able to increase the performance of the model.

8 Acknowledgment

The author would like to thank eSmart Systems for the support in the work with this paper.

References

[1] Z. Akata, M. Malinowski, M. Fritz, and B. Schiele. Multi-cue zero-shot learning with strong supervision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 59–68, 2016.

[2] P. M. Burlina, A. C. Schmidt, and I. Wang. Zero shot deep learning from semantic attributes. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pages 871–876, 2015.

[3] LianLi Gao, Jingkuan Song, Junming Shao, Xiaofeng Zhu, and HengTao Shen. Zero-shot image categorization by image correlation exploration. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, ICMR '15, page 487–490, New York, NY, USA, 2015. Association for Computing Machinery.

[4] Jingcai Guo and Song Guo. A novel perspective to zero-shot learning: Towards an alignment of manifold structures via semantic feature expansion. *ArXiv*, abs/2004.14795, 2020.

[5] Li Liu Ling Shao Haofeng Zhang, Yang Long. Adversarial unseen visual feature synthesis for zero-shot learning. *Neurocomputing*, 329(7):12–20, 2019.

[6] N. Karessli, Z. Akata, B. Schiele, and A. Bulling. Gaze embeddings for zero-shot image classification. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6412–6421, 2017.

[7] C. H. Lampert, H. Nickisch, and S. Harmeling. Attribute-based classification for zero-shot visual object categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):453–465, 2014.

[8] Y. Liu and T. Tuytelaars. A deep multi-modal explanation model for zero-shot learning. *IEEE Transactions on Image Processing*, 29:4788–4803, 2020.

[9] J. Markowitz, A. C. Schmidt, P. M. Burlina, and I. Wang. Hierarchical zero-shot classification with convolutional neural network features and semantic attribute learning. In *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*, pages 194–197, 2017.

[10] Jenssen R. Roverso D. Nguyen, V.N. Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning. *International Journal of Electrical Power and Energy Systems, 99, 107-120*, 2018.

[11] F. Nian, Y. Sheng, J. Wang, and T. Li. Zero-shot visual recognition via semantic attention-based compare network. *IEEE Access*, 8:26002–26011, 2020.

[12] T. Lillicrap koray kavukcuoglu O. Vinyals, C. Blundell and D. Wierstra. *Matching Networks for One Shot Learning*. D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Eds.Curran Associates, Inc., 2016, pp. 3630–3638.

[13] P. Peng, Y. Tian, T. Xiang, Y. Wang, M. Pontil, and T. Huang. Joint semantic and latent attribute modelling for cross-class transfer learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(7):1625–1638, 2018.

[14] J. Qin, Y. Wang, L. Liu, J. Chen, and L. Shao. Beyond semantic attributes: Discrete latent attributes learning for zero-shot recognition. *IEEE Signal Processing Letters*, 23(11):1667–1671, 2016.

[15] Christopher D. Manning Andrew Y. Ng Richard Socher, Milind Ganjoo. Zero-shot learning through cross-modal transfer. 2013.

[16] F. Shen, X. Zhou, J. Yu, Y. Yang, L. Liu, and H. T. Shen. Scalable zero-shot learning via binary visual-semantic embeddings. *IEEE Transactions on Image Processing*, 28(7):3662–3674, 2019.

[17] K. Wang, S. Wu, G. Gao, Q. Zhou, and X. Jing. Learning autoencoder of attribute constraint for zero-shot classification. In *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 605–610, 2017.

[18] X. Xu, T. Hospedales, and S. Gong. Semantic embedding space for zero-shot action recognition. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 63–67, 2015.