Exploring Latent Class Structures in Classification-By-Components Networks

Lars Holdijk
Radboud University Nijmegen
larsholdijk@gmail.com

Abstract

A Classification-by-Components network (CBC) operates under the assumption that every input image can be classified based on its decomposition into a set of components. An important characteristic of these components is that they can be used in the decomposition of images from different classes. The components are class independent. In this work, we discuss the latent class structure encoded in the sharing of components between classes. We propose to visualize this structure using a Shared Component Graph (SCG). Consecutively we discuss the insight into the decision making process of a CBC the visualization can provide.

1. Introduction

A Classification-By-Components network (CBC) [4] classifies its input by structurally decomposing it based on a set of learned components. The underlying assumptions are inspired by the theory that humans recognize complex objects in a similar manner, introduced by I. Biederman and coined "Recognition-by-Components" [1]. An important criteria in this theory is that the components on which the classification is based are class independent, allowing components to be used in the recognition of multiple objects.

An example of this would be the classification of the classes passenger car and sports car, two distinct classes in the ImageNet [3] dataset. Both classes have individual components, *e.g.* seats for the passenger car and a spoiler for the sports car, but they also share a number of components, such as wheels and the hood of the car.

Where the class specific components obviously play a crucial role in making the distinction between the two cars, the shared components are equally important for the dataset wide classification problem. For this reason, CBCs learn class independent components and reason over all of them to classify the input image. Basically, a CBC learns for each component how its presence provides evidence in favour or against the image being of a specific class.

In this work, we look at the relationship between classes learned by a CBC by investigating their shared components. First we will however provide a short introduction to CBCs and consecutively discuss how the shared component naturally encode a class structure. By visualizing the latent class structure of CBCs using a novel Shared Component Graphs (SCG), we both qualitatively and quantitatively evaluate the capability of CBCs to recover the relationship between classes. We find that CBCs can uncover intuitive relationship between classes in a self-supervised manner. Additionally, SCGs show to be an important tool for finding flaws in the classification process of component based classifiers.

2. Background: Classification-By-Components

Given an input image $\mathbf{x} \in \mathbb{R}^{n \times n}$, a CBC first decomposes it based on a set of components $\mathbf{k} \in \mathbb{R}^{m \times m}$, where $n \geq m$. It does so using a *detection probability function* $d_{\mathbf{k}}(\mathbf{x})$, resulting in the probability for each component to be present in the image. The detection probability function is implemented using a *sliding distance measure* that calculates the similarity between a patch of the input $\mathbf{x}$ and a component $\mathbf{k}$. By processing the input $\mathbf{x}$ using a feature extractor f_θ before computing the similarity between $f_\theta(\mathbf{x})$ and $\mathbf{k}$ the detection probability function can be improved. The components $\mathbf{k}$ as well as the parameters θ of the feature extractor are trainable parameters of the classifier.

Given the detection probability $d_{\mathbf{k}}(\mathbf{x})$ for every input $\mathbf{x}$ a reasoning process is applied to determine the final classification decision. The detection of each component can either contribute positively, negatively or not at all to the probability of the input belonging to class c. The probability of the component contributing positively, negatively or not at all is represented in the reasoning probabilities of a class c for component $\mathbf{k}$, denoted respectively as $r^+_{c,\mathbf{k}}, r^-_{c,\mathbf{k}}$ and $r^0_{c,\mathbf{k}}$. Combining the detection probability and the reasoning probability, the contribution of component $\mathbf{k}$ to the probability of the input belonging to class c is calculated as

$$p_{c,\mathbf{k}}(\mathbf{x}) = \frac{d_{\mathbf{k}}(\mathbf{x}) \cdot r^+_{c,\mathbf{k}} + (1 - d_{\mathbf{k}}(\mathbf{x})) \cdot r^-_{c,\mathbf{k}}}{(1 - r^0_{c,\mathbf{k}})}. \qquad (1)$$

By summing these probabilities for all components the final class hypothesis probability is achieved. Similar to the

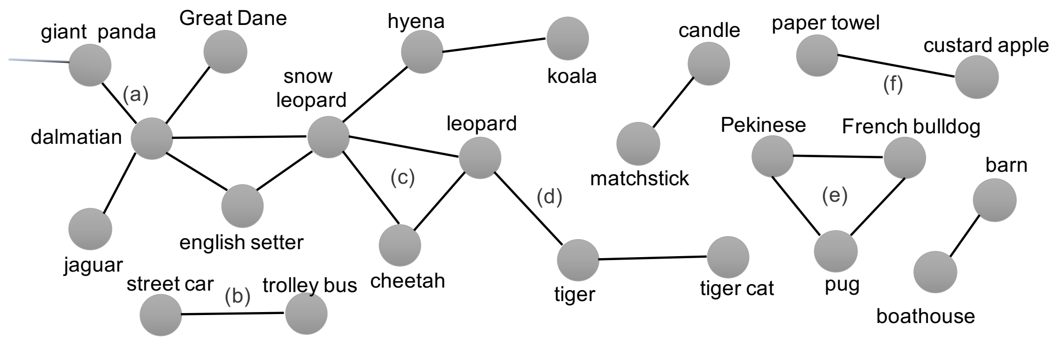

Figure 1: Shared Component Graph visualization of the latent class structure learned by the CBC discussed in [4]

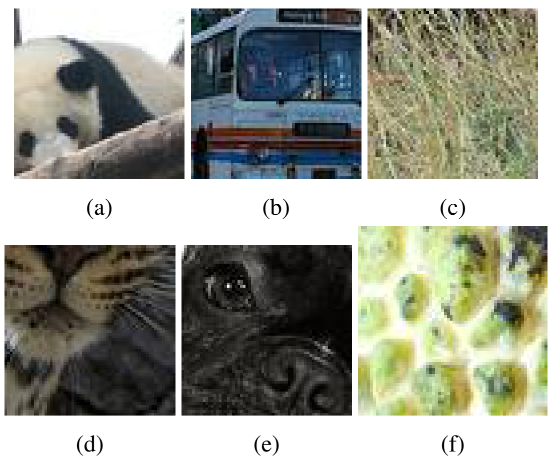

Figure 2: A selection of shared components represented in the SCG in Fig. 1

components, the reasoning probabilities are trainable parameters of the classifiers.

3. Shared component Graph

The components learned by a CBC are class-independent, meaning that they play a role in the classification process of all classes. A component $\mathbf{k}$ can contribute positively to the class hypothesis probability of class c_i, by having a high $r^+_{c_i,\mathbf{k}}$, and negatively to class c_j, by having a high $r^-_{c_j,\mathbf{k}}$. Similarly, classes c_i and c_j could consider the presence of the same component important. In this case both $r^+_{c_i,\mathbf{k}}$ and $r^+_{c_j,\mathbf{k}}$ are high. The aforementioned role of the wheel component in the classification of the classes passenger car and sports car is a clear example of this. In essence, two classes with a high positive reasoning probability for the same component can be said to *share the component*. Assuming that a shared component between two classes is an indication of the classes being considered similar by the CBC, the collection of classes sharing components can be considered its latent class structure. We propose Shared Component Graphs (SCG) to visualize this.

A SCG is a (disconnected) graph where each node represents a class and edges signal that the two connected classes share one or more components. Formally an edge between classes c_i and c_j is drawn when both $r^+_{c_i,\mathbf{k}} > t$ and $r^+_{c_j,\mathbf{k}} > t$ for any component $\mathbf{k}$. Where threshold $t \in [0, 1]$ is chosen to represent the certainty that the presence of the component is important for the classes.

4. Case study 1: ImageNet

In the following we will take a look at the latent class structure learned by the CBC trained on ImageNet by Saralajew et al. [4] . In their introductory paper Saralajew et al. trained a CBC with a ResNet50 feature extractor to classify the 1000 classes contained in the ImageNet dataset using 5000 components. We will use a SCG to visualize the class structure learned by the CBC and discuss it for a subset of the classes. Following this, we will use the WordNet hierarchy of the ImageNet dataset to quantitatively evaluate the capability of CBCs to recover this hierarchy. Note that this is achieved in a self-supervised manner. During training of the CBC only class labels were provided, without any indication of the relation between the classes.

In Fig. 1 the SCG of the CBC trained by Saralajew et al. on the ImageNet database is visualized. To create the SCG a threshold of 0.8 was used. The SCG of the ImageNet CBC is a disconnected graph with a number of connected *clusters* (in graph theory known as components, for clarity we will use the term cluster). Note that only a small subset of the full SCG is given. This subset is chosen to be illustrative of the expressive capability of the SCG. A large number of single nodes, small clusters (2-4 nodes) and larger cluster (5+ nodes) were omitted. The full SCG can be found at `http://larsholdijk.com/scg.html`. In Fig. 2 a subset of the components belonging to the edges in the SCG are given. The relation between components and edges is given by the in brackets enclosed letters. The components were extracted from the CBC as described in [4].

The largest cluster in the SCG contains a total of 33 nodes, of which a small sample is featured on the left of Fig. 1. The full cluster extends from the giant-panda class onwards. All nodes within the cluster represent classes originating from the animals synset in the ImageNet database. Within the cluster a large number of black-and-white animals can be found, such as the snow leopard, English setter, Great Dane, giant panda and dalmatian. The latter of which, connected by component (a) in Fig. 2, were also discussed in [4]. In addition to being black-and-white some of these animals also feature a dotted pattern. This dotted pattern can also be found in different colours on animals such as jaguars, hyenas and cheetahs, also found in the cluster.

The disconnected graph, however, mainly consist of smaller clusters, representing shared components between two or three classes. Thematically, these smaller cluster still contain related classes, such as street cars and trolley buses, barns and boathouses or candles and matchsticks. The same can be said for the fully connected cluster Pekinese, French bulldog and pug that in fact share the same component (e). This is also the case for the two sets of fully connected subclusters of three nodes within the larger cluster. This shows that CBCs are not only capable of uncovering class relations between two classes but also between larger groups of classes. However, there are also a number of unrelated classes that share components. In Fig. 1 this is highlighted by the class paper towel and custard apple sharing component (f). While component (f) closely resembles a custard apple, it does not relate intuitively to a paper towel.

In total, the SCG of the CBC trained on ImageNet contains 156 edges representing shared components, connecting together 198 classes. The strongest connection can be found between the classes indigo bunting and jay, both birds. The indigo bird is also the most connected class, sharing 14 components with other classes.

4.1. ImageNet WordNet

In contrast to many others, classes in the ImageNet dataset are organized using a hierarchy called WordNet. This makes ImageNet an excellent dataset for quantitatively evaluating the latent class structure of a CBC. Using the WordNet tree as a representation of the true class structure of ImageNet, we can validate the latent class structure encoded in the shared components of a CBC.

For this purpose, we define the true distance between two classes to be the shortest path of one of the two classes to a common ancestor in the ImageNet WordNet tree. For example, the distance between the class leopard and snow leopard equals 1 as they are siblings within the tree with the big cat synset as parent. The distance between leopard and hyena however equals 2, as their closest common ancestor carnivore is two steps away from hyena (hyena - canine - carnivore) and three steps from leopard (leopard - big cat - feline - carnivore). This distance definition limits the impact of difference in verbosity between sub-trees. With a distance metric in place we can assign a quality score to each shared component, or edge in the graph.

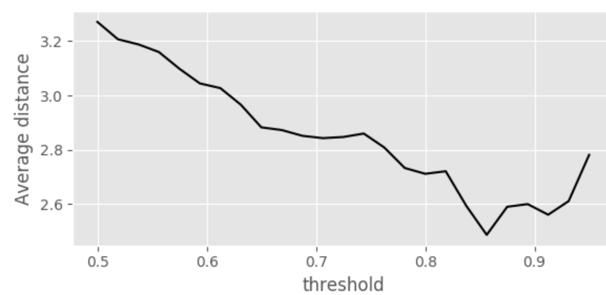

Figure 3: Influence of the chosen threshold on the average distance between two classes sharing a component.

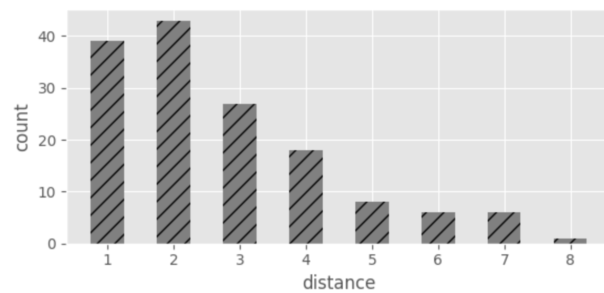

Figure 4: Histogram of the distance between classes in the recovered WordNet hierarchy by a CBC.

Its obvious that with increasing the threshold of acceptance, the number of shared components found is reduced. Ideally however, the class combinations with small distance between them should have a shared component with a higher shared positive reasoning probability, maintaining the shared component status with higher thresholds, then classes that are far apart in the WordNet hierarchy. In Fig. 3 the average distance between classes that share a component is given for different thresholds. The graph shows that this desired behaviour is the case for the ImageNet CBC. At a threshold of 0.86 the minimal average distance between classes that share a component is found. At this threshold, on average the classes that share a component have a distance of 2.5 in the WordNet hierarchy.

In Fig. 4 a histogram of the assigned scores is given for a SCG with threshold 0.8. Over half of all shared components are shared by classes with an ancestor within 2 steps. The

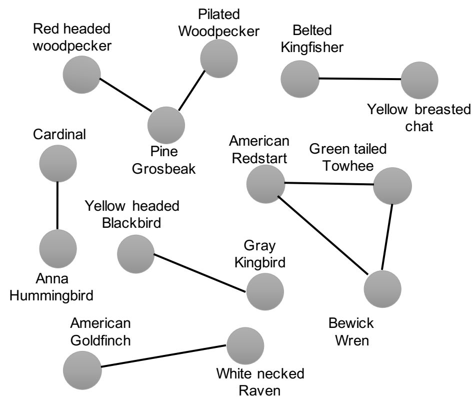

Figure 5: SCG of a CBC trained on the CUB dataset.

number of shared components with a larger distance drops after this. We see this as preferred behaviour. While the CBC only recovers a small portion of all class relationships, it does recover the relationship between similar classes.

5. Case study 2: Bird species classification

Another interesting dataset for exploring the latent class structure in CBCs is the CUB-200-2011 (CUB) dataset [5]. The CUB dataset contains a total of 11788 images distributed over 200 bird species. While no direct class hierarchy is provided, the relation between classes is obvious due to the taxonomy of bird species.

We trained a CBC using a DenseNet with 121 layers as feature extractor to classify the images in the CUB dataset. The CBC contained 1000 components of size $64 \times 64 \times 3$ and used the cosine similarity in combination with global maxpooling to achieve a single detection probability for each component. To align our work with that of [4], the same training procedure was used to retain the components in the feature space and reduce complexity. The network proposed in [2] by Chen et al. is very similar to CBCs, and hence serves as a good benchmark. Similar to their work, the DenseNet was pretrained using the ImageNet dataset. [1]

The trained CBC achieved a validation accuracy of 82.7%, outperforming the ProtoPNet of Chen et al. In Fig. 5 the full SCG representing the latent class structure of the trained CBC is given for a threshold of 0.65. In contrast to the SCG of the CBC trained on ImageNet only a small portion of the classes is connected through a shared component and no large clusters of classes can be found. This is despite the significantly lower threshold. Similarly, no strong thematic connection can be found between the classes that share components, aside from the three classes on the top left. All of these classes represent red birds. With the threshold set lower, more classes share components. However, as we know from the evaluation using the ImageNet WordNet hierarchy, this will also reduce the semantic significance of the shared components.

Ultimately, the taxonomy of bird species tells us that there is a relation between the classes in the CUB dataset. The fact that, despite its high accuracy, the CBC did not uncover this structure shows that validation accuracy alone is not sufficient for evaluating a component based classifier.

[1] Unfortunately, the test set of the CUB dataset does contain some images found in the ImageNet training dataset. Due to the architecture of CBCs, which only used the DenseNet as feature extractor, this does not influence the final classification layers.

6. Conclusion

The usage of a SCG for visualizing the latent class structure learned by a CBC is a powerful method for extending their already existing interpretability. Using the ImageNet dataset and its WordNet hierarchy, we showed that using the SCG we could gain a better understanding of how the CBC arrived at its final classification decision. For example, it heavily relies on the use of colour and low level pattern matching. In addition to this, it also aided in finding current flaws within the decision making process. The unexpected shared component between the classes paper towel and custard apple is an example of this. Similarly, the SCG of the CUB dataset CBC showed that SCGs can uncover flaws that classic metrics such as validation accuracy can not.

In future work it is important to extend the notion of shared components to other component based classifiers, such as [2], and feature representation in normal neural networks. However, these methods often do not follow the strict probability assertion on the reasoning weights of CBCs that make the threshold possible. This hurdle can potentially be overcome using softmax on the trained weights.

References

[1] Irving Biederman. Recognition-by-components: a theory of human image understanding. *Psychological review*, 94(2):115, 1987.

[2] Chaofan Chen, Oscar Li, Daniel Tao, Alina Barnett, Cynthia Rudin, and Jonathan K Su. This looks like that: deep learning for interpretable image recognition. In *Advances in Neural Information Processing Systems*, pages 8928–8939, 2019.

[3] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[4] Sascha Saralajew, Lars Holdijk, Maike Rees, Ebubekir Asan, and Thomas Villmann. Classification-by-components: Probabilistic modeling of reasoning over a set of components. In *Advances in Neural Information Processing Systems*, pages 2788–2799, 2019.

[5] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011.