

On the Relationship between Visual Attributes and Convolutional Networks (Supplementary Material)

Victor Escorcia^{1,2}, Juan Carlos Niebles², Bernard Ghanem¹

¹King Abdullah University of Science and Technology (KAUST), Saudi Arabia

²Universidad del Norte, Colombia

1. Conv-net details

In the experiments of Sections 3.2 and 3.3, we use activations from all convolutional and fully-connected layers of the *Alexnet* architecture. In this case without loss of generality, we decouple the output layer in two pieces: (i) a fully-connected layer performing a linear combination of *fc-7* outputs, called *fc-8*, and (ii) a soft-max operation which uses *fc-8* as inputs and produces the posteriors associated with each object class. On the other hand, we only use activations from all convolutional layers, *fc-6* and *fc-7* of *Alexnet* architecture for our ablation study (Section 3.4). The underlying reason is to avoid the undesirable effect of directly degrading the classification performance due to the ablation of nodes from the output layer.

2. Ablation Study: Complementary Results

Ablating Individual Attributes: In Section 3.4, we ablated ACNs corresponding to all 25 attributes by taking the union of their sparse supports. In that experiment, we see that ablating ACNs significantly degrades overall object recognition performance. However, we expect a similar behavior when ablating ACNs corresponding to individual attributes separately. We report these results for 4 attributes in Figure 1. As would be expected, the drop-off in top-5 accuracy is less drastic than that shown in Figure 8 of the submission. This is due to the fact that the percentage of ablated nodes is much less. Figure 1 shows this degradation for the ILSVRC-12 validation set, when 4 attributes ('black', 'furry', 'rectangular', and 'spotted') are ablated from the conv-net separately. These attributes represent a sample attribute from the four attribute groups that were defined for the ImageNet-Attribute dataset.

Ablating Sets of Attributes: Figure 2 shows more qualitative results of our ablation study (Figure 9 in the submission) and the impact of ACNs on object classification. We observe that the conv-net seems to make use of attributes from the image context to recognize some objects, such as 'Ping-Pong Ball' or 'Puck' associated with the 'Green, Brown, Wooden' attribute set or 'Valley' associated with 'Black, Brown, Furry'.

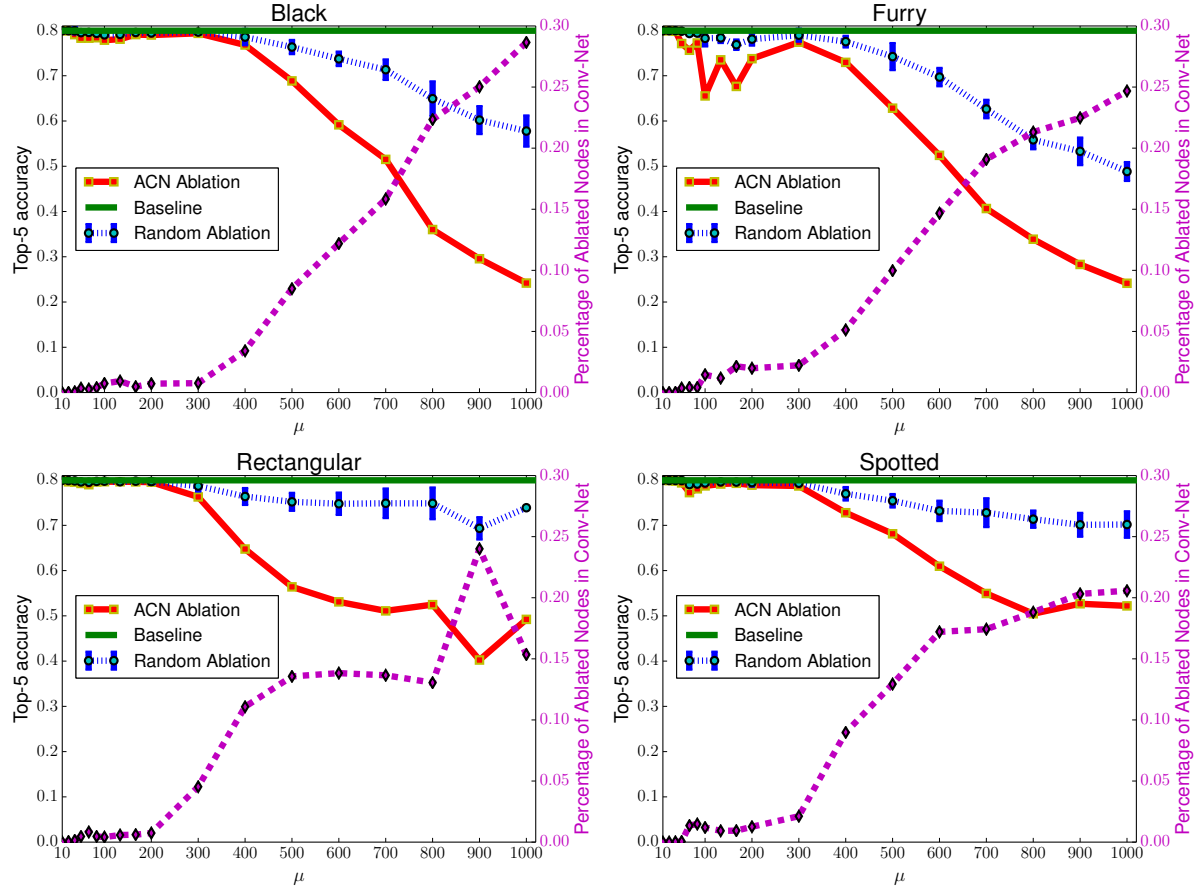


Figure 1. Plots top-5 accuracy on the ILSVRC-2012 validation set of the original *Alexnet* model (green) and its two surgically damaged variants. One variant (red) ablates the ACNs of one attributes (at each μ value), while the other (blue) ablates an equal number of randomly sampled nodes. The ablated attribute is on top of each graph. Both variants show a steep drop-off as μ increases; however, the difference in accuracy between the two is significant. This suggests that ACNs encode important information used by the conv-net for recognition





































































































Black Brown Furry										
	Indri (-0.50)	Chesapeake Bay Retriever (-0.38)	Gazelle (-0.33)	Irish Setter (-0.33)	Red Wolf (-0.32)	Orangutan (-0.31)	Vizsla (-0.30)	Dhole (-0.30)	Valley (-0.29)	Labrador Retriever (-0.29)
										
Green Red Orange Yellow										
	Butternut Squash (-0.50)	Pomegranate (-0.42)	Acorn (-0.36)	Spaghetti Squash (-0.36)	Banana (-0.31)	Acorn Squash (-0.31)	Orange (-0.29)	Guacamole (-0.26)	Strawberry (-0.26)	Beer Glass (-0.24)
										
Gray Metallic Shiny										
	Crane (-0.63)	Police Van (-0.58)	Yawl (-0.51)	Bobsled (-0.51)	Pickup (-0.50)	Banjo (-0.50)	Wreck (-0.50)	Barrow (-0.50)	Lawn mower (-0.48)	Racer (-0.48)
										
Furry Gray White										
	Ruffed Grouse (-0.49)	Arctic Fox (-0.48)	French Bulldog (-0.46)	Hen-of-the- woods (-0.45)	Cardigan (-0.44)	Pomeranian (-0.43)	Samoyed (-0.43)	Brittany Spaniel (-0.42)	Baboon (-0.42)	Guenon (-0.41)
										
Green Brown Wooden										
	Vulture (-0.75)	Broccoli (-0.74)	Pool Table (-0.71)	Crane (-0.69)	Tennis Ball (-0.69)	Ping-Pong Ball (-0.69)	Puck (-0.68)	Black Stork (-0.66)	Nematode (-0.66)	Screw (-0.65)
										
	Sunglass (-0.08)	Sunglasses (-0.08)	Drill Power (-0.07)	Flute (-0.07)	Toilet Tissue (-0.07)	Spatula (-0.07)	Sunscreen (-0.06)	Night Snake (-0.03)	Appenzeller (-0.01)	Velvet (-0.01)

Figure 2. Shows object classes that are the most (red box) and least (green box) affected by ablating ACNs corresponding to five example attribute groups. The mean average precision degradation of each of these classes is reported below its representative image. The most affected classes tend to contain the ablated attributes, while the least affected ones do not.