Attempts to Interpret a Neural Network

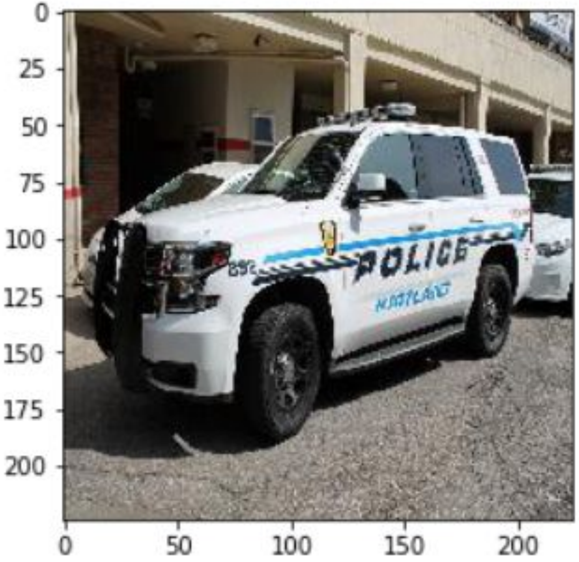
Attempts to Interpret a Neural Network

We can

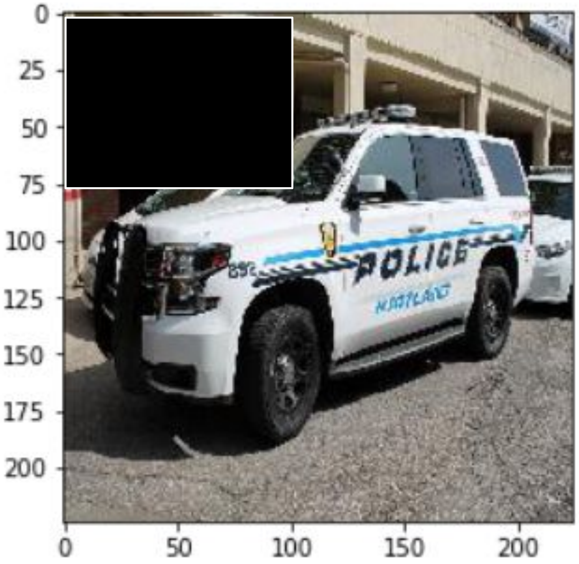
- Understand the model architecture
- Visualize the filters / weights
- Extract the output of intermediate neurons / layers
- **Locate important parts of the image according to the model**

Attempt 4 - Locate important parts of the image

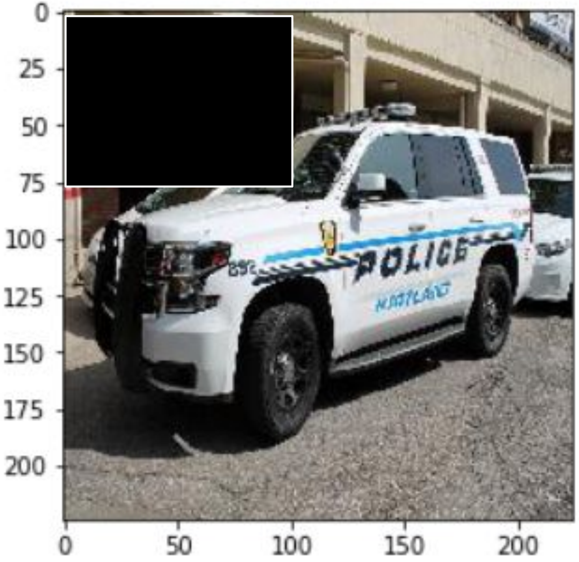
Attempt 4 - Occlusion Maps



Attempt 4 - Occlusion Maps

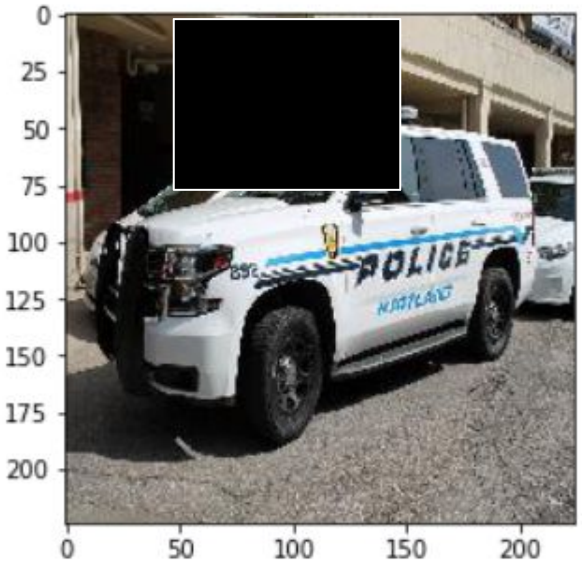


Attempt 4 - Occlusion Maps



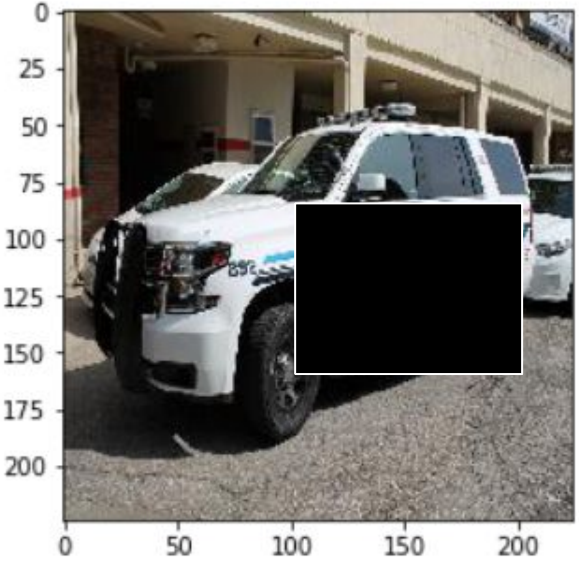
95% confident

Attempt 4 - Occlusion Maps



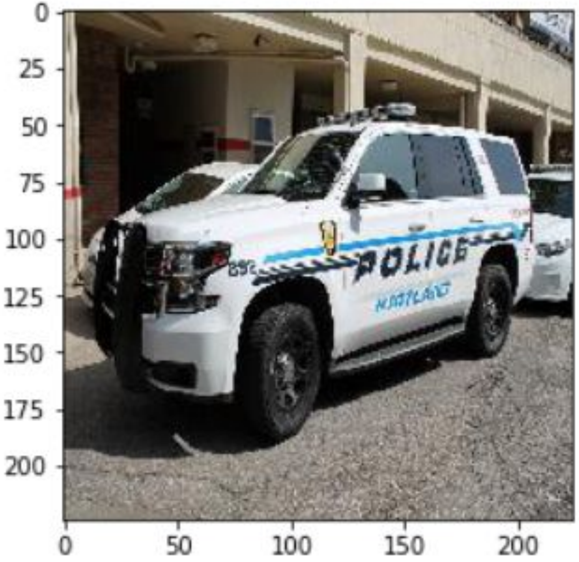
90% confident

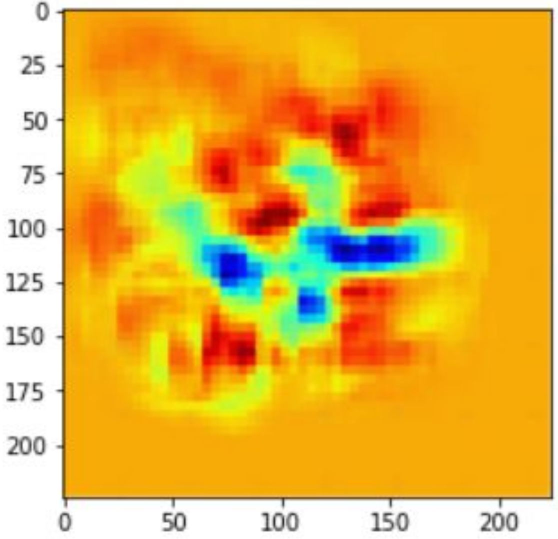
Attempt 4 - Occlusion Maps



20% confident

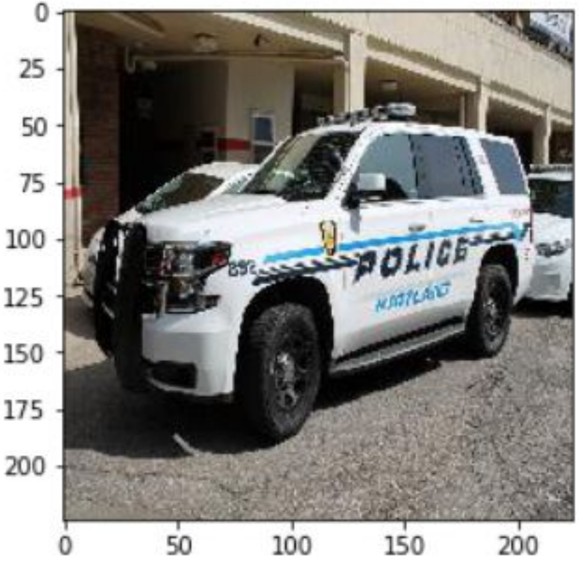
Attempt 4 - Occlusion Maps



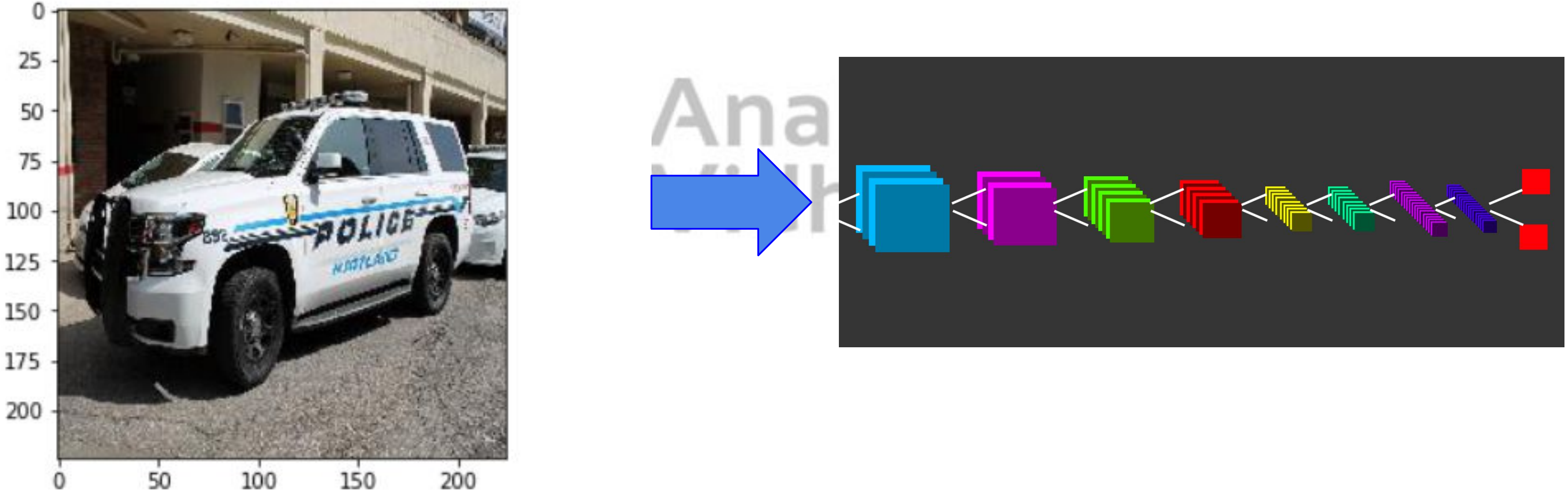


Analytics Vidhya
Learn everything about analytics

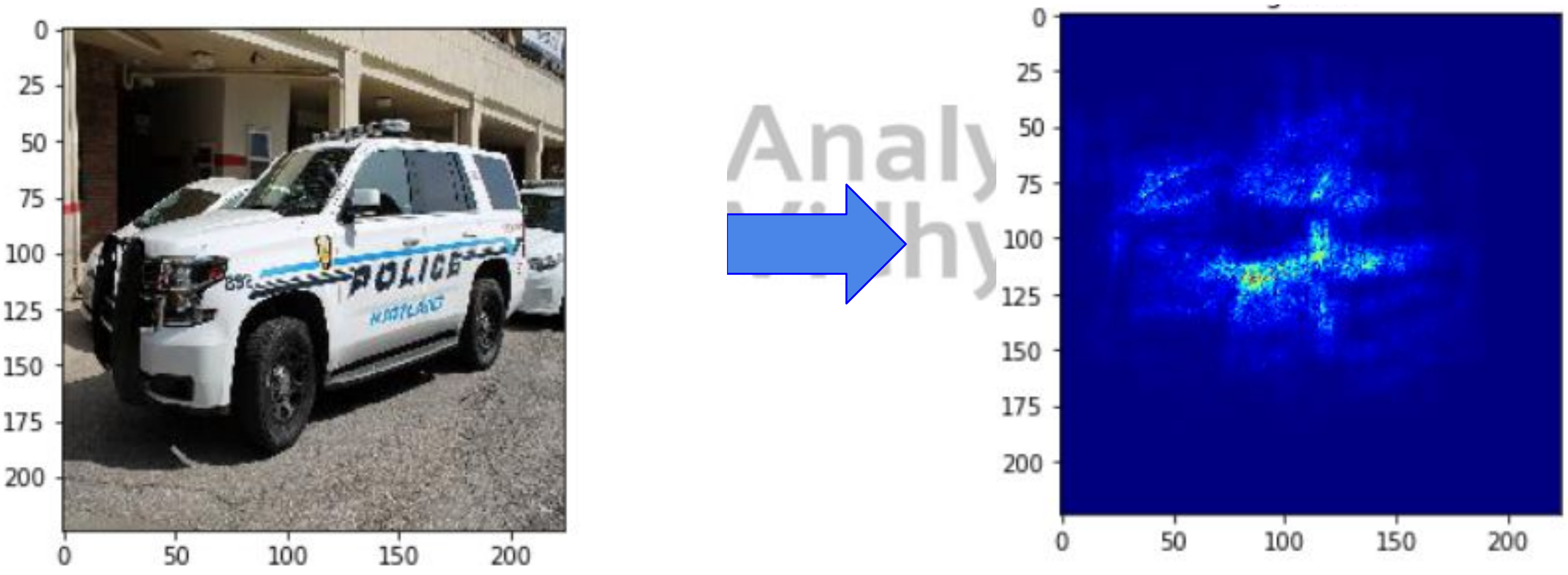
Attempt 4 - Saliency Maps



Attempt 4 - Saliency Maps



Attempt 4 - Locate important parts of the image



Attempts to Interpret a Neural Network

We can

- Understand the model architecture
- Visualize the filters / weights
- Extract the output of intermediate neurons / layers
- Locate important parts of the image according to the model

Thank You