Laboratory

Interpretability

Deep Learning for Artificial Intelligence (DLAI)

Oscar Pina
PhD Student
Universitat Politècnica de Catalunya
oscar.pina@upc.edu

Acknowledgements

Amaia Salvador
PhD 2019
Universitat Politècnica de Catalunya

Daniel Fojo
MSc 2019
Universitat Politècnica de Catalunya

Xavier Giro-i-Nieto
Associate Professor
Universitat Politecnica de Catalunya

@DocXavi
xavier.giro@upc.edu

Albert Mosella
Ph.D. Student
Universitat Politècnica de Catalunya

In today's lab...

We are going to see how to:

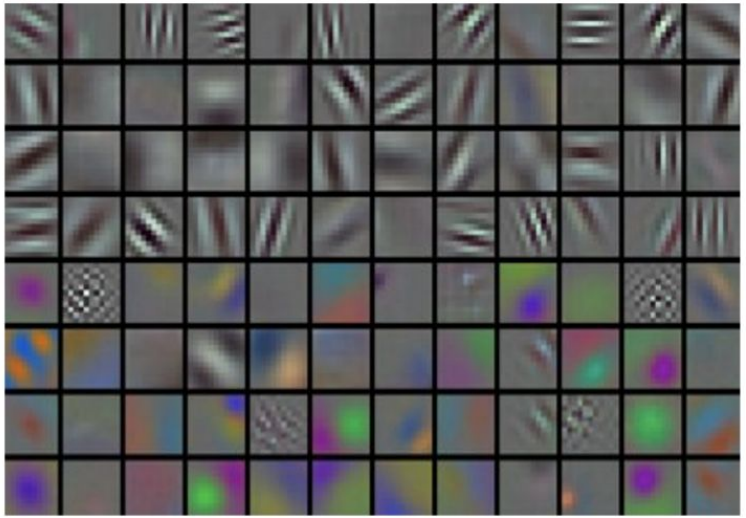
- Visualize filters
- Find the top k samples that activates a neuron
- Occlusion sensitivity experiment
- T-SNE
- Synthesizing images to maximize activations

Credits: SLazebnik

In today's lab...

We are going to see how to:

- **Visualize filters**
- Find the top k samples that activates a neuron
- Occlusion sensitivity experiment
- T-SNE
- Synthesizing images to maximize activations

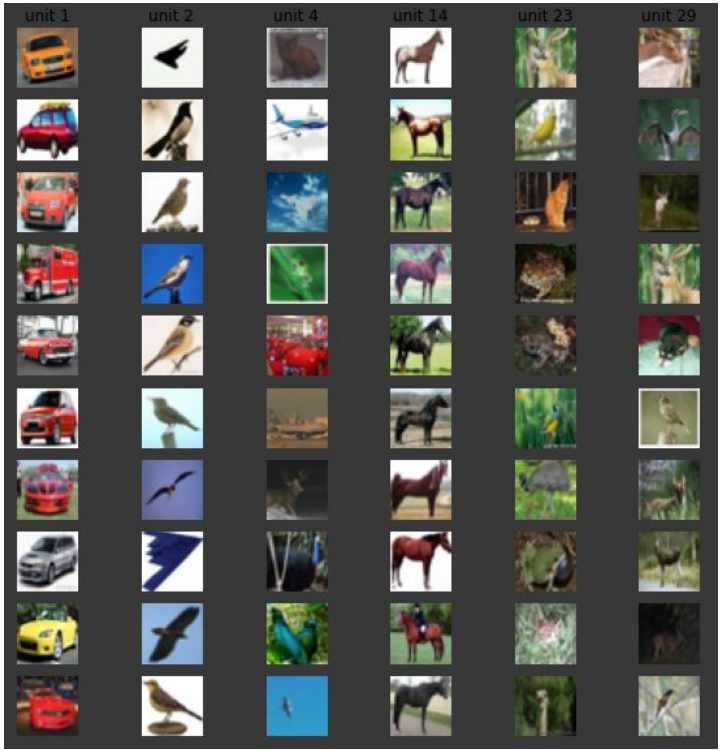


Credits: SLazebnik

In today's lab...

We are going to see how to:

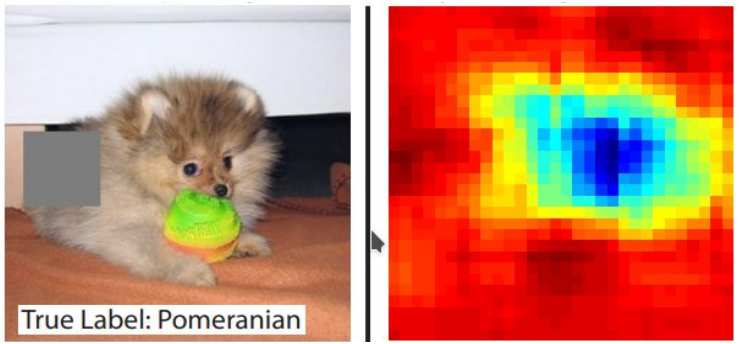
- Visualize filters
- **Find the top k samples that activates a neuron**
- Occlusion sensitivity experiment
- T-SNE
- Synthesizing images to maximize activations



In today's lab...

We are going to see how to:

- Visualize filters
- Find the top k samples that activates a neuron
- **Occlusion sensitivity experiment**
- T-SNE
- Synthesizing images to maximize activations

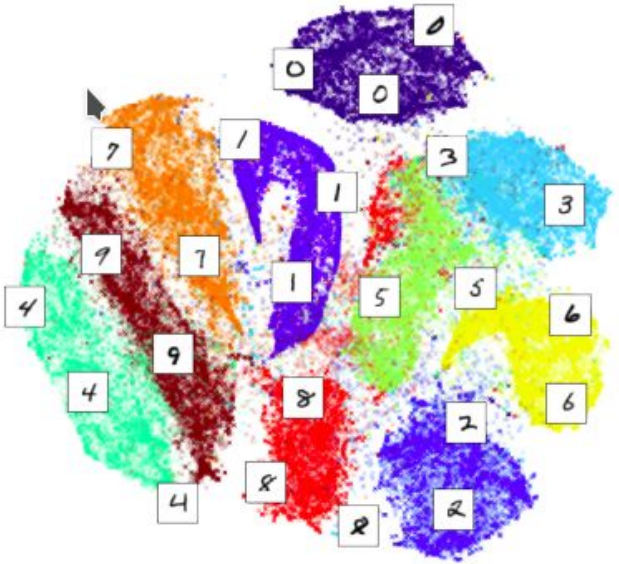


Credits: M. Zeiler and R. Fergus, Visualizing and Understanding Convolutional Networks, ECCV 2014

In today's lab...

We are going to see how to:

- Visualize filters
- Find the top k samples that activates a neuron
- Occlusion sensitivity experiment
- **T-SNE**
- Synthesizing images to maximize activations



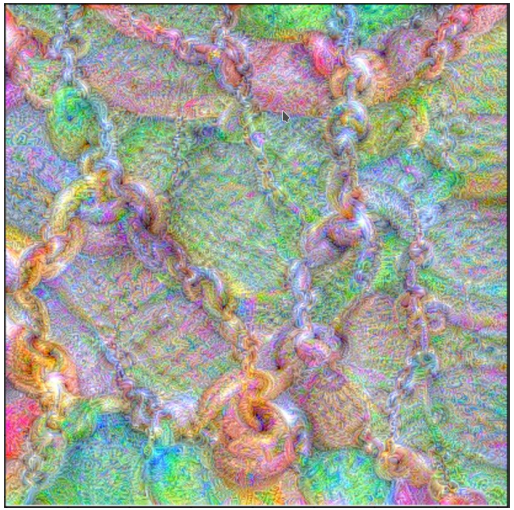
tSNE on MNIST digits

Credits: Maaten, Laurens van der, and Geoffrey Hinton. "Visualizing data using t-SNE." Journal of machine learning

In today's lab...

We are going to see how to:

- Visualize filters
- Find the top k samples that activates a neuron
- Occlusion sensitivity experiment
- T-SNE
- **Synthesizing images to maximize activations**



Kick off the lab

1. Launch a web browser (Chrome recommended).
2. Login with your @estudiantat.upc.edu GSuite account.
3. Create a copy the notebook of this lab to your Gdrive.
4. (Right) Click on the file and choose Open File with “Google Colaboratory”