

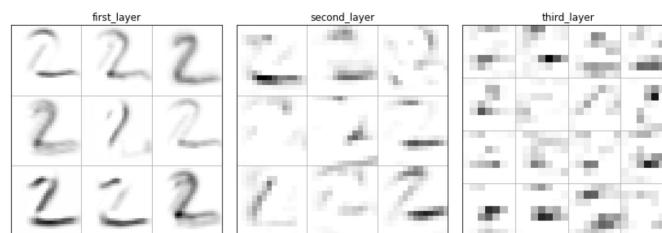
PW-10

{dorian.magnin, noemien.kocher@master.hes-so.ch}

Present the results obtained in point 1 and comment those results. E.g., describe the patterns that are shared by different digits and that are being used by the CNN to solve the recognition task. Can you infer what do the activation maps of the L2 and L3 layers represent ?

Ce qu'on remarque, c'est que plus on avance dans les couches, plus il est difficile (pour un humain en tout cas) de déduire des informations pertinentes.

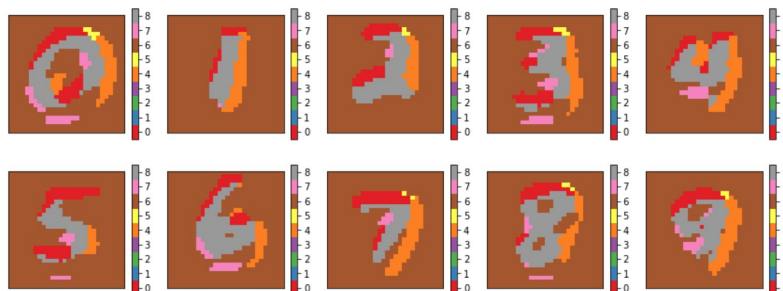
Par exemple:



On remarque bien les contours sur la première et deuxième couche, mais il est quasi impossible de déduire quoique ce soit sur la dernière couche.

Les activations moyennes nous permettent de confirmer que c'est le cas pour tous les digits.

Visualize which filter is activated most of the time for each pixel



Le 0 met en évidence des parties du cercle.

Le 1 n'est en pratique pas utilisé. Le 1 est représenté comme une partie de 7 et 4.

Le 2 et le 3 ne sont pas présents non plus. Ils sont représentés comme parties d'autres chiffres.

Le 4 est représenté comme une barre verticale à droite. Ce qui se comprend, vu que c'est une caractéristique unique du 4.

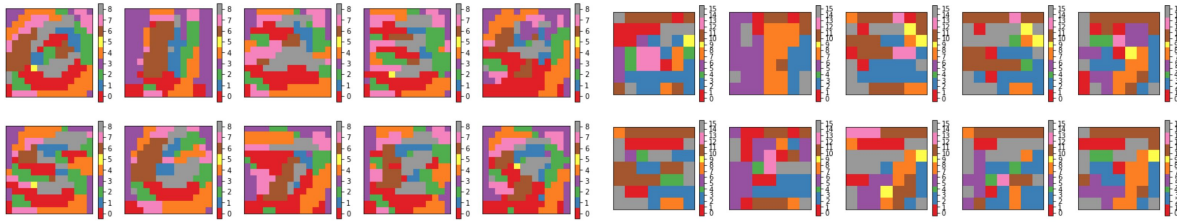
Le 5 n'est pas très présent, voire presque inutile.

Le 6 est invisible sur nos graphiques (brun sur brun).

Le 7 est un peu comme le 5, peut utiliser.

Le 8 est étonnamment énormément utilisé pour représenter tous les chiffres et on ne retrouve pas vraiment sa forme.

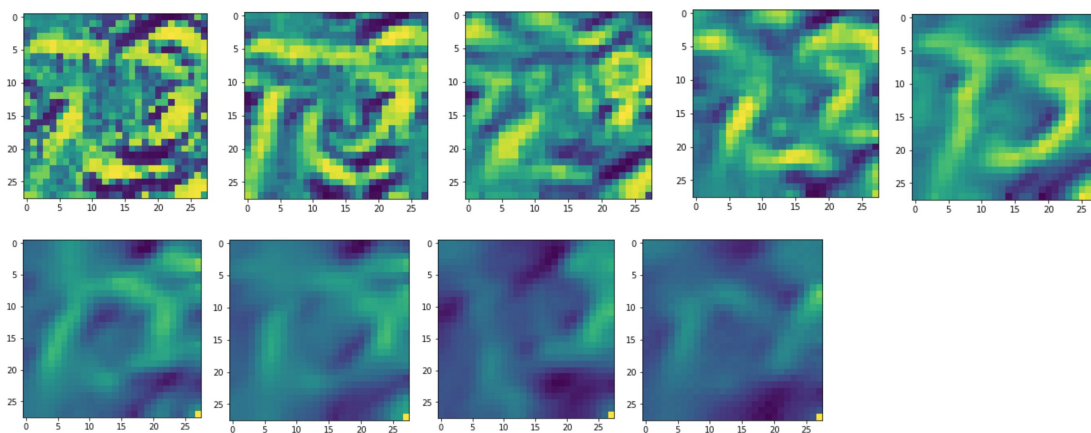
La suite des représentations nous offre de magnifiques représentations artistiques:



Present the results of the experiments proposed at the end of the notebook corresponding to “activity maximization” and answer the question concerning the use of such a technique.

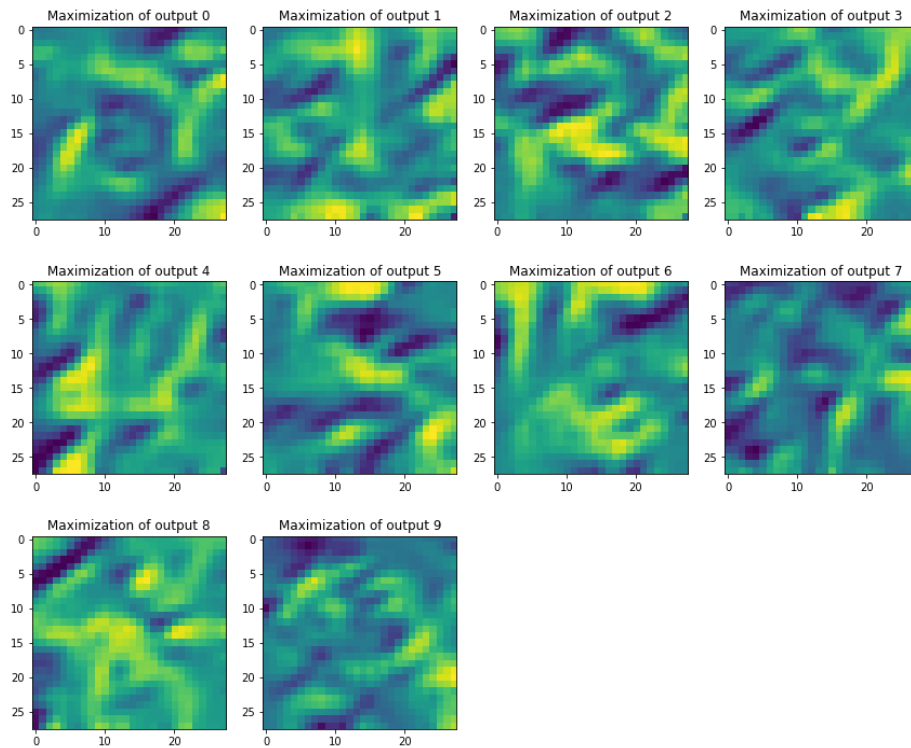
- Test different values of `tv_weight`
 - Try values between 0.1 and 20 (for example: [0.125, 0.25, 0.5, 1, 2, 4, 8, 16])
 - Select the regularization parameter that gives the best images (more realistic)
 - Show the images that maximize each one of the outputs of the network
- Maximize two outputs at the same time (`filter_indices=[f1, f2]`)
 - Try two classes with similar shape like 1 and 7 or 4 and 9
 - Try two classes with very different shapes like 0 and 1 or 7 and 8
 - How activation maximization can be useful for understanding a deep neural network? Explain

Voici les résultats, avec `tv_weight` (dans l'ordre) = [0.125, 0.25, 0.5, 1, 2, 4, 8, 16]



On va utiliser `tv_weight = 1`

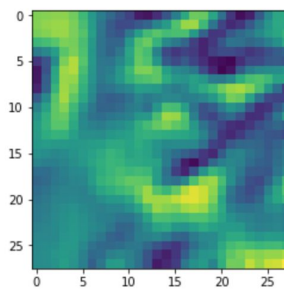
Show the images that maximize each one of the outputs of the network



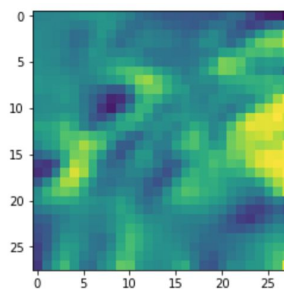
On remarque vaguement les formes des chiffres, mais c'est assez flou. Le 4 se voit très bien.

- Maximize two outputs at the same time (`filter_indices=[f1, f2]`)
 - Try two classes with similar shape like 1 and 7 or 4 and 9
 - Try two classes with very different shapes like 0 and 1 or 7 and 8
 - How activation maximization can be useful for understanding a deep neural network? Explain

Pour le 5 et le 6 (ils sont similaires):



Pour le 7 et le 0 (sont différents):



How activation maximization can be useful for understanding a deep neural network?

Explain

En comprenant comme le système “réfléchis” et se trompe, on peut du coup essayer d’éviter au maximum que le système se trompe en faisant du preprocessing sur certaines images.

On peut réagir au “hacking” du système en le rendant plus intelligents.

On peut remarquer que certain filtres sont inutiles et les enlever. De manière générale, comprendre comment un système fonctionne permet de l’optimiser.