Hes·so | MSE | MASTER OF SCIENCE IN ENGINEERING

MACHINE LEARNING (T-MACHLE)

Practical work 11 - Understanding Deep Neural Networks

STUDENT: ***ROMAIN CLARET***
PROFESSOR: ***A. PEREZ-URIBE & J. HENNEBERT***
ASSISTANT: ***H. SATIZABAL***
DUE-DATE: ***MONDAY 10 DECEMBER 2018***

PRESENT AND COMMENT THE RESULTS OBTAINED IN POINT 1

Describe the patterns that are shared by different digits and that are being used by the CNN to solve the recognition task.

- I can notice a similar abstraction progression at each layer, for each digit. The first 2 layers, for me human, are interpretable mainly because it contains spatial information. However, starting from the 4th layer, each layer is using internal abstraction based on the previous network layer, which encodes the data for us, human.

Can you infer what do the activation maps of the L2 and L3 layers represent?

- From what I am understanding, L2 and L3 are extracting the meaningful spatial features.

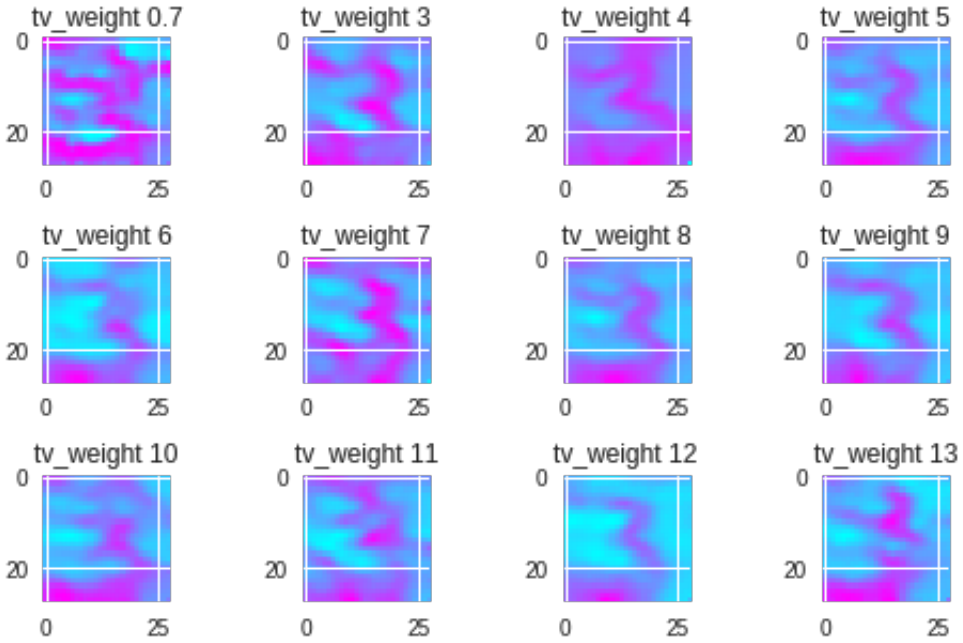
PRESENT THE RESULTS AND ANSWER THE QUESTION FOR EACH TECHNIQUE.

Activity maximization

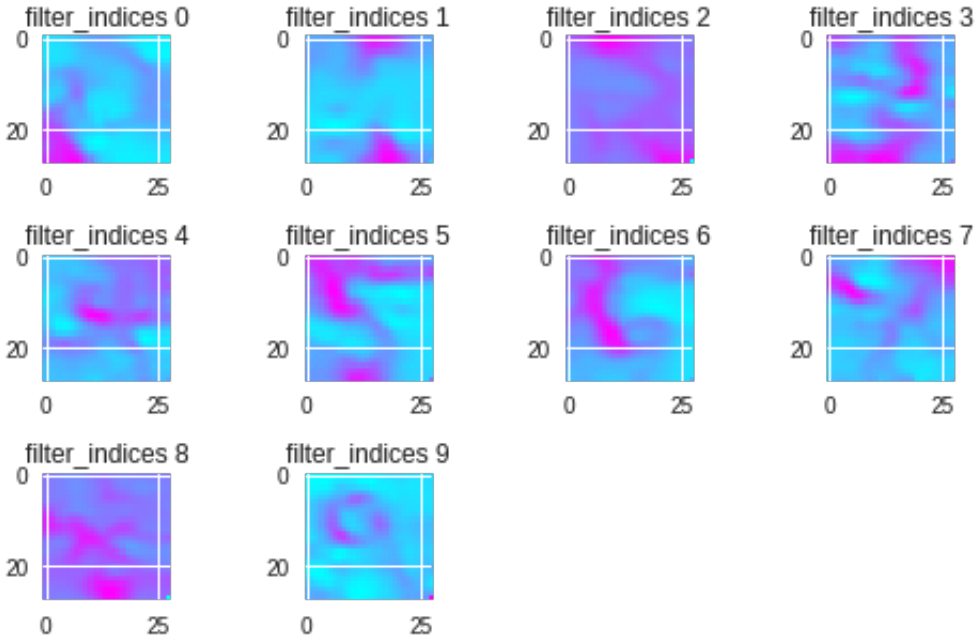
Test different values of tv_weight

- I used `filter_indices=3`, it was too hard for me to find something realistic for `filter_indices=0`, and I used contrasting colors to help me too.

- Understood that `tv_wight` is linked to the `figsize` and `subplot`.
- I choose subjectively the more realistic regularization parameter **12** from the list:
  `[0.7, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13]`

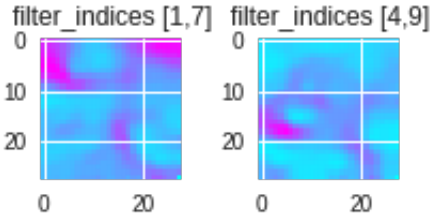


- Just for fun, what each element looks like

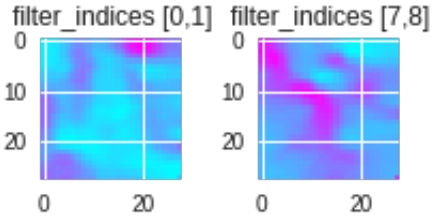


Maximize two outputs at the same time (filter_indices=[f1, f2])

For two classes with similar shape like 1 and 7 and 4 or 9



For two classes with very different shapes like 0 and 1 or 7 and 8



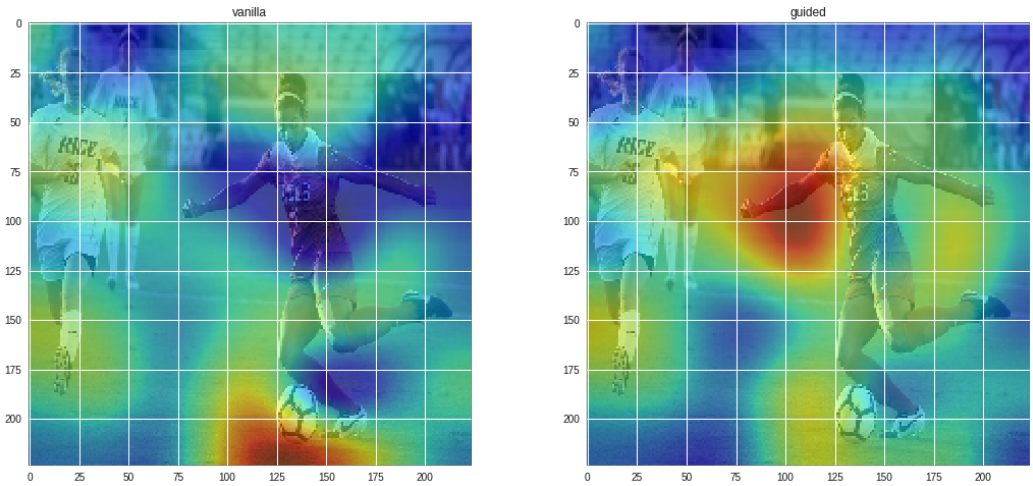
How activation maximization can be useful for understanding a deep neural network? Explain

- It could be helpful to optimize raw data with preprocessing, target more specific features, or even combine classes.

Class Activation Maps

soccer03.jpg

Does the map reflect the class of the pictures?



- Top 5 predictions:

CLASSES	PROBABILITIES

soccer_ball	0.65820456
racket	0.09809169
ballplayer	0.087732404
baseball	0.043914024
rugby_ball	0.042742964

- The top 1 class is reflecting well the presence of a soccer ball, found with the vanilla mapping.
- The rest of the class are probably suggested based on the guided mapping, which is mainly targeting the arm of the soccer player.

Are there images in which the activation map highlight zones that do not belong to the object? Which ones?

- Yes, even if the probabilities are much lower, the rest of the predicted classed are suggesting a baseball game instead of a soccer game. It's targeting the players, arms, heads, and legs. For me, it's suggesting a motion.

Hypothesize about the behavior above.

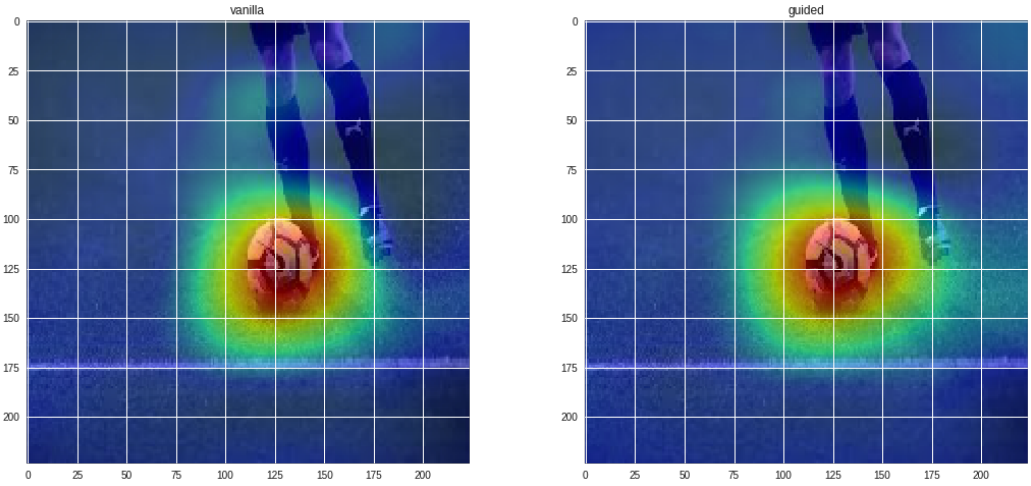
- It actually makes sense, due to the position of the soccer player which is very similar to what we would expect from a ball player.

Based on ImageNet, are images have been labeled appropriately? explain.

- For the soccer_ball, it was indeed correctly labeled.
- However, for the second best class **Racket**, it's a mistake, probably due to the arm targeting.
- Also, for the rest, based on the keyword **soccer** it's a mistake since we can find similar pictures.

soccer02.jpg

Does the map reflect the class of the pictures?



- Top 5 predictions:

CLASSES	PROBABILITIES
soccer_ball	0.9329394
football_helmet	0.04838616
rugby_ball	0.006925001
ballplayer	0.0045021377
baseball	0.004369618

- Yes.

Are there images in which the activation map highlight zones that do not belong to the object? Which ones?

- Not really.

Hypothesize about the behavior above.

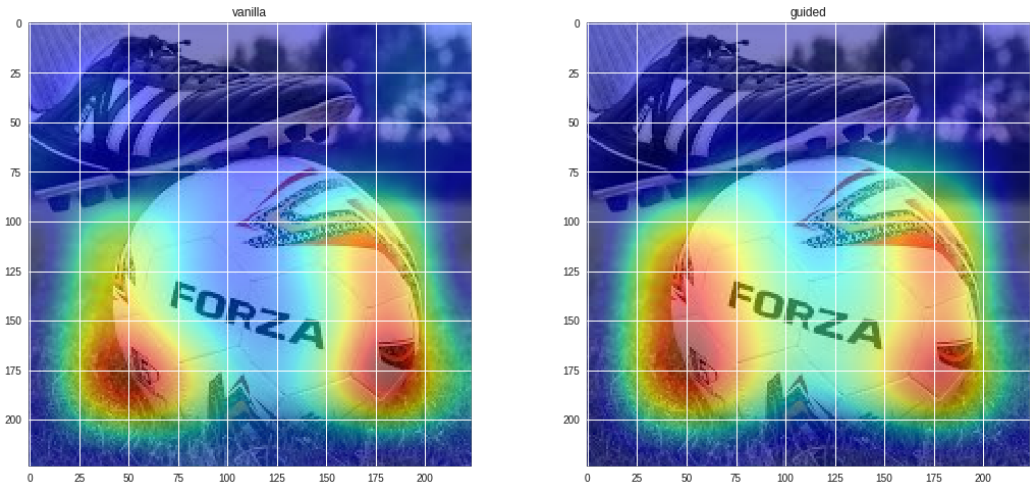
- The vanilla and guided mapping did a perfect job at targeting the ball.

Based on ImageNet, are images have been labeled appropriately? explain.

- Yes, because it was a proper targeting.

soccer01.jpg

Does the map reflect the class of the pictures?



- Top 5 predictions:

CLASSES	PROBABILITIES
soccer_ball	0.6119669
rugby_ball	0.38721398
volleyball	0.00078869733
football_helmet	2.0883166e–05
baseball	5.2446503e–06

- It's good, the top 1, is correct.
- However, the second prediction is incorrect, but the ball abstraction is present.

Are there images in which the activation map highlight zones that do not belong to the object? Which ones?

- Not really, it's targeting the edges of the ball.

Hypothesize about the behavior above.

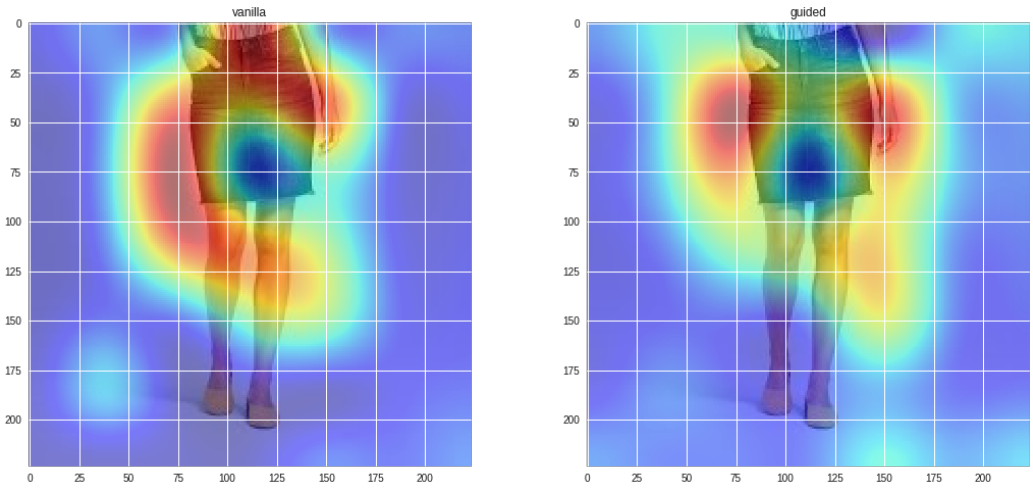
- Since it's targeting the ball edges instead of the ball itself. The probability for any kind of balls is making sense.

Based on ImageNet, are images have been labeled appropriately? explain.

- The rugby_ball shape is making sense based on the 2 edges activated.
- The soccer_ball is correct.

miniskirt.jpg

Does the map reflect the class of the pictures?



- Top 5 predictions:

CLASSES	PROBABILITIES
miniskirt	0.5854138
jean	0.18462898
velvet	0.057723965
overskirt	0.02316156
maillot	0.020552302

- Yes it does. All classes are making sense based on both mappings.

Are there images in which the activation map highlight zones that do not belong to the object? Which ones?

- The mapping is quite good.

Hypothesize about the behavior above.

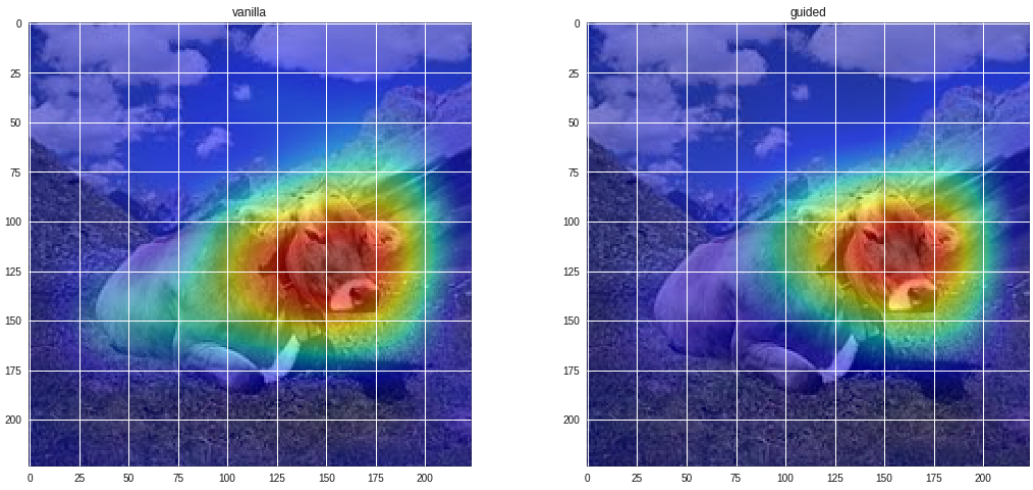
- It got the naked legs, hands positions, and skirt edges. It's impressive, based on the top 3, it describes well the object: Miniskirt, in jeans texture with a velvet effect on it.

Based on ImageNet, are images have been labeled appropriately? explain.

- Yes, note that often the women in miniskirt are holding their hands the same way as on the picture.

cow.jpg

Does the map reflect the class of the pictures?



- Top 5 predictions:

CLASSES	PROBABILITIES
ox	0.43208572
ram	0.1094391
bighorn	0.06092967
ibex	0.052729726
water_buffalo	0.04500427

- Yes, the predictions are based on the head.

Are there images in which the activation map highlight zones that do not belong to the object? Which ones?

- No.

Hypothesize about the behavior above.

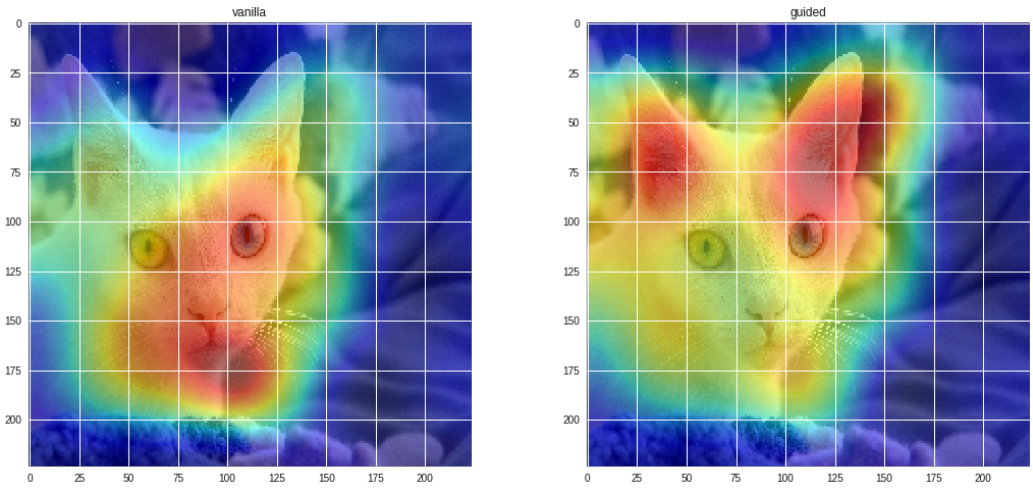
- It's targeting the heard. It's difficult to know the sex of beef, and it has horns.

Based on ImageNet, are images have been labeled appropriately? explain.

- I couldn't find pictures of **ox**, but for the cow label, we often see heads, and it's the same for the rest of the top 5.

cat01.jpg

Does the map reflect the class of the pictures?



- Top 5 predictions:

CLASSES	PROBABILITIES
Egyptian_cat	0.92138076
tiger_cat	0.017774599
tabby	0.017314434
lynx	0.012366519
Siamese_cat	0.004968628

- Yes, again, it's the head that is targeted.

Are there images in which the activation map highlight zones that do not belong to the object? Which ones?

- No

Hypothesize about the behavior above.

- The targeting and the class are making sense.

Based on ImageNet, are images have been labeled appropriately? explain.

- Yes, when checking Egyptian_cat, we can also find tiger_cat inside…
- Heads are also very present.