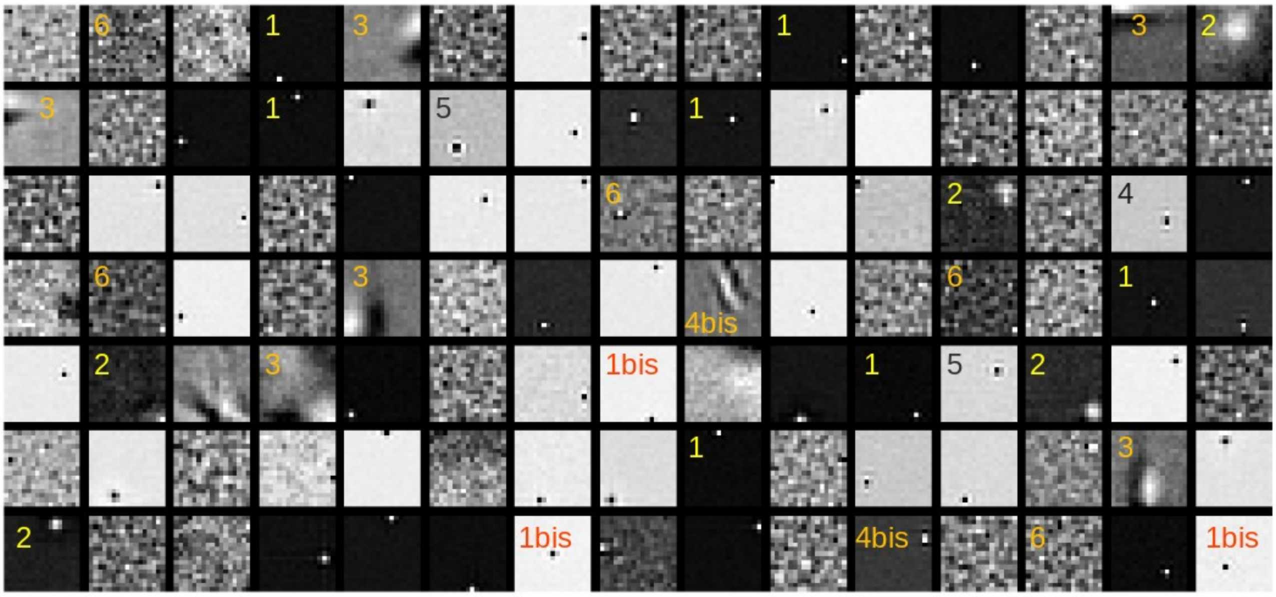
IP Paris & Univ. Paris Saclay

Masters Data AI / AI

Image Mining Exam – December 2023

1 – Convolution Kernels



The following figure shows a sample of 105 neurons from the first layer of a deep network that has been trained for image denoising. Each neuron has 17x17 weights so that its activation value corresponds to the output of a square convolution kernel applied to each location of the image. The corresponding convolution kernels are represented hereunder using greylevels.

Try to categorise those neurons according to the different types of convolution kernels, and for each category, explain the effect on the image and the mathematical interpretation of the result.

The neurons can roughly be grouped into 6 categories (examples are labelled in the image above), as follows:

1) The neurons of type (1), and their opposite (1bis) correspond to singleton convolution kernel that, when applied to the image, produces a simple translation (or an inverted version of the translation for 1bis neurons). Mathematically those neurons simply apply a projection of the neighbourhood (patch) vectors into the canonical basis.

2) The neurons of type (2) are just smoothed version of the previous ones, their application on the image will produced blurred translated images.

3) The neurons of type (3) are first derivative kernels. Note the different scales (size of the black and white blobs) and phases (their positions). Their application on the image will enhance/detect edges whose orientation is orthogonal to the orientation of the black and white blobs.

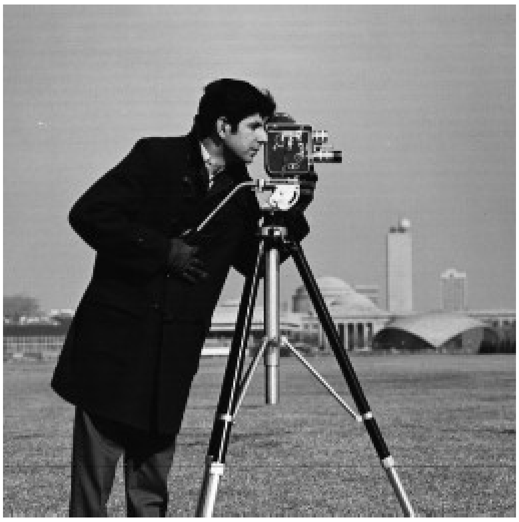
4) The neurons of type (4) (or their opposite versions 4bis) are second derivative kernels. Their application on the image will enhance/detect curvatures whose orientation is orthogonal to the orientation of the black and white blobs.

5) The neurons of type (5) are Laplacian kernels. Their application on the image will enhance the edges in the image, whatever their orientations (isotropic property of the Laplacian).

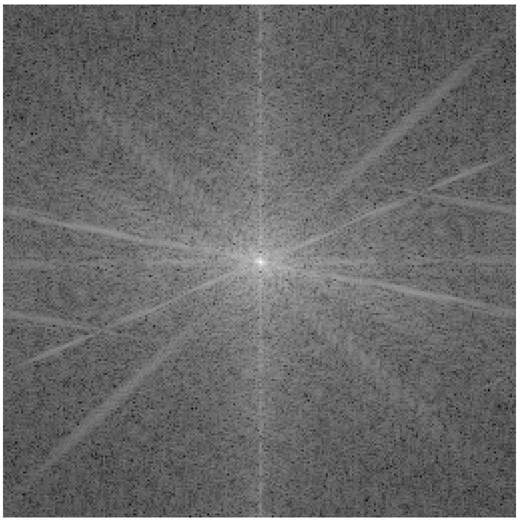
6) The remainder neurons (6) are difficult to interpret; they may be used to extract complex features that are not understandable at the first layer level, or more prosaically they may be neurons whose weights have not converged and thus remained to their initial random states.

2 – Fourier transform

The following figure displays a grayscale image and its Fourier amplitude spectrum.



(a)



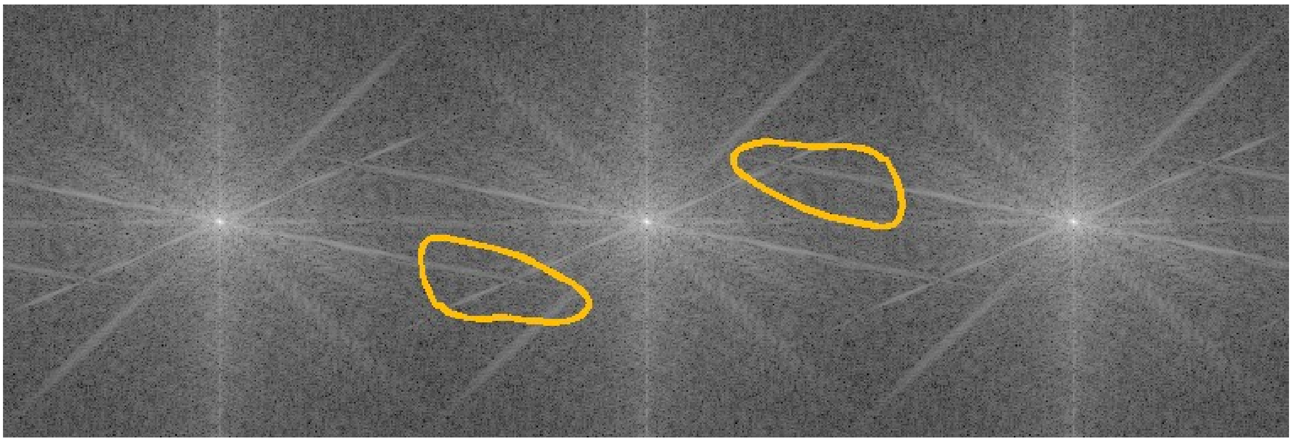
(b)

Try to interpret the spectrum (b) in relation to the structures present in image (a). Something is visible from the spectrum that is hard to see in the image, can you find what it is? Finally, mention two practical applications of the Fourier transform in Computer Vision.

1. The amplitude spectrum shows bright lines, that correspond in the spatial domain to high contrast linear structures in the orthogonal direction; this is related to the decomposition of a beam of parallel lines into its fundamental and harmonics, that will appear as aligned points in the Fourier domain.

2. Overlapping parts are clearly visible at the middle left and right of the amplitude spectrum, which correspond to the continuity of one of the main diagonal lines by periodic repeating of the spectrum (see Figure below). This indicates an aliasing problem in the image domain, due to a bad sub-sampling, albeit not visible in the image.

3. Applications of the Fourier transform: Texture analysis, Calculation of large (e.g. Gaussian) convolution kernels, Calculation of correlation filters in object tracking, Image registration,...



3 – Feature spaces for Content Based Image Retrieval

What are, in general, the expected properties of a feature space that should be used to represent an image, in a task of Content-Based Image Retrieval? Give 2 examples of possible feature spaces.

In general, a feature space used to represent an image is expected to (1) faithfully represent the types of categories that will be found in the queries (2) be robust enough to the alterations and variability that can affect the images in the corresponding categories, and (3) be calculable efficiently. Here are some examples of possible feature spaces:

- *Bag of features representation, i.e. distribution-based measures (e.g. histogram) of quantised local descriptors coming from handcrafted or deep features, and sparsely or densely calculated over the image.*
- *Metrics over a latent space learned by a deep network, typically pre-trained on a classification task.*

If the closest images of some query image are those minimising a distance in the aforementioned feature space, what precautions should be taken in designing the distance depending on the feature space?

The distance must respect the structure / geometry of the feature space, e.g. Euclidean space, distribution space, Riemannian space...

Considering a data base made of a very large set of images collected from different internet platforms like Flickr, Pinterest and Tumblr, without any annotation, timestamp or meta-data. **What feature space would you propose in order to retrieve images for the following kinds of query:**

1. **photo of a sea shore**
2. **photo of a horse**
3. **photo of Donald Trump**
4. **photo a Van Gogh's painting**

The solutions for this question are not unique, but here are examples of relevant feature spaces for the different cases:

1. *Sea shore: should represent the layout of the image as well as dominant colours or textures; a bag-of-features approach made with a codebook on deep features can make it.*
2. *Horse: a latent space extracted from a deep network trained on classifying horse image seems a good choice.*
3. *Donald Trump: the use of a face detector, associated with the latent space of a deep network designed for face recognition and trained with Trump's faces is advised.*
4. *Van Gogh: should represent the style, which can be thought as the stationary (textural) part of the image, that can be modelled from global statistics like: Fourier descriptors, Covariance matrices of deep features, etc.*

4 – Object Tracking

Considering the problem of Tracking an arbitrary object, i.e. an object that has not been learned by a detector. Is **there a part of the tracking algorithm that can be learned *offline* (i.e. before the object to be tracked is known)? If yes, describe such part.**

Yes, at least two aspects of tracking can be learned offline:

1. *The feature space, which is used to calculate the matching metrics; this can be taken from the latent space of a deep network trained in a similar environment.*
2. *The ability to match an object template to the location of a query image where the most similar object lies; this can be learned from e.g. a Siamese network.*

Now describe 3 different methods to learn *online* (i.e. during the run time) a model of the tracked object, or at least to adapt the algorithm to changes in the object appearance and behaviour.

1. *Learning a correlation filter by incrementally minimising a loss based on the Fourier transform to obtain a target correlation map.*
2. *Update recursively a histogram of some features for distribution based (e.g. Mean Shift) tracking.*
3. *Update smoothly a local descriptor representation (e.g. R-Table) for a part based (e.g. Hough) tracking.*

What are the difficulties in designing a good evaluation metrics to assess a tracking algorithm?

This problem is particularly difficult; the basic principle consists in summing/averaging over the sequence duration, some localisation accuracy measure between the tracking support (e.g. bounding box) and the ground truth position. However two main difficulties arise:

1. *Time integration: an algorithm that loses the target at the end of the sequence is favoured with respect to an algorithm that loses it during the first frames, which does not necessarily reflects the right comparison. This is why the current evaluation protocols provide - after a certain penality frames - a reset to the ground truth position to resume the tracking evaluation.*
2. *Localisation accuracy: current matching measures like Jaccard index (ratio of intersection over union areas between candidate and ground truth bounding boxes) strongly disadvantages small targets. This is why it may be recommendable to use soft metrics like sum over distance maps.*

5 – Deep Neural Networks

1. **What is the primary purpose of an activation function in a neural network?**
2. **What is the main function of backpropagation in neural network training? Write – in pseudo-code – the training script.**
3. **What does dropout aim to prevent in a neural network?**
4. **What is the role of the learning rate in training a neural network?**
5. **What is the advantage of using transfer learning in deep learning?**

1- The primary role of activation functions is to introduce non-linearity in neural networks in order to prevent them from acting as mere linear regressor/separator.

2- The function of backpropagation is to update the weights of the neurons in a gradient descent process that minimizes the loss function. Its pseudocode can be written as follows:

```
For (x_batch, y_batch) in Dataloader:
    # Forward
    y_predict = Model(x_batch)
    Loss = loss(y_batch,y_predict)
    # Backpropagation
    Grad = Model.backward(Loss)
    Optimization_step(Model,Grad)
```

3- Dropout aims to avoid overfitting to the training data, by preventing any group of neurons from taking on major importance in the model, thus introducing diversity that favours robustness.

4- The role of the learning rate is to control the step size in weight updates, defined by the relative magnitude of the gradient vector in the gradient descent performed by backpropagation.

5- Transfer learning seeks to accelerate the convergence of training a model A by taking advantage of another model B trained on a different domain or a different task, which is expected to be more beneficial to A than being trained from scratch.

6 – Remote Sensing

You are working in a large railway company. To improve your maintenance scheme, you want an exhaustive map of your train network that highlight the track structures that need special maintenance such as bridges and the vegetation surrounding the tracks that need to be regularly pruned.
You are conducting a study to choose the best sensors for this task, and the associated AI framework.
After looking online, you find these characteristics for the sensors:

	Optical sensors		SAR sensors	
	Sentinel-2	Pléiades	Sentinel-1	TerraSAR-X/TanDEM-X
Resolution	10m	0.70m	10m	1m
Swath	290km	14km	250km	10km
Repeat time	5 days	25 days	6 days	11 days
3D	No	Tri-stereo mode	Interferometry with acquisitions at two different days	-Interferometry with acquisitions at two different days - Same day interferometry with both satellites
Cost	0€	5000€ per image	0€	5000€ per image

1- Knowing that a train track is approximately 1.5m wide, which optical or SAR sensors can you use to detect the train tracks?
Given the width of the train track, high resolution data such as Pléiades and TerraSAR-X, that have a resolution below 1m, are needed.

2-

a) Amongst: train tracks, bridges and vegetation, which element(s) would you detect using SAR and which element(s) would you detect using optical data?
Vegetation is best-detected using optical sensor thanks to the infrared band, from which the NDVI can be computed. The appearance of bridges and tracks can change a lot. Bridges are often composed of metallic structures, which can be easily detected on SAR images. For the tracks, it depends on their orientation. Oriented parallel to the sensor trajectory, they will be brighter than oriented perpendicular since a double bounce will be created with the side of the tracks.

b) Do you think that the gain in detection performances of using two types of images will be worth the price of acquiring these two types of images?
This question is difficult and open. Given the price of the images, and the number of images needed to train NN, it might be enough to start with SAR images only. Moreover, vegetation areas that are larger than train tracks may be detected on Sentinel-2 images that are free. However, it is very important to have close acquisition dates for the SAR images and the optical images. Asking for SAR acquisition only at the start might hinder the gain of using both SAR and optical afterward.

3- If you want to use a deep-learning framework, a training database will be needed. Thanks to IGN open data, you have airborne optical images of France acquired in spring every 4 years, sampled at

0.5m, with the annotation of train tracks. Which problems will you face if you want to use this database to train your network to detect train tracks on:

a) optical satellite data ? Which training procedure can you use to mitigate these problems?

Since the gap in resolution will be small, working on under sampled airborne images may be enough to get good performances.

b) winter optical satellite data? Which training procedure can you use to mitigate these problems?

Shadow and vegetation will have a different appearance. Transferring the annotation of train tracks on winter optical data to fine tune the network may be needed. Otherwise colorimetric data augmentation may also work.

c) optical satellite data over India? Which training procedure can you use to mitigate these problems?

The appearance of vegetation and of the train tracks may change a lot. Colorimetric data augmentation will be needed. Semi supervised learning that builds on non-annotated data will be useful as well.

7 – Aesthetics and Artificial Intelligence

1 - Image Beauty assessment is a sub-problem of Image Quality assessment.

True ○ False ●

2 - Assessing beauty of a photo using purely objectivist criteria is no better than chance (around 50% success).

True ○ False ● (around 80 % correct classification in 2 classes with AVA database)

3 - When using handcrafted primitives and classification methods for aesthetic assessment of images, photo expert primitives (rule of third, power of the center, vanishing background,…) provide no better results than computer vision primitives (SIFT, SURF, wavelet coefficients,…).

True ● False ○

4 - A frequently used method for subjectivist assessment of beauty, makes use of OCEAN Method. OCEAN stands for “Optimal, Coherent, Exact, Adaptive, Neutral”.

True ○ False ● (Openness, Conscientiousness, Extroversion, Agreeableness, Neuroticism)

5 – Recommendation techniques, which are widely used in mail order to suggest new products to a customer, are very effective in assessing the judgement of a user exposed to a new photograph.

True ○ False ●

6 – A rather efficient technique of subjective assessment of image beauty for a given user is to fine-tune a network trained with AVA database, with two final layers which are trained with the user’s photos as posted on his/her personal website.

True ● False ○

7 – Education and culture are much more important to contribute to the aesthetic assessment of photos than user’s temper, mood and contextual circumstances of the judgement experiment.

True ○ False ●

8 – Short networks, with convolutional layers only, have been proven more efficient than deep neural networks to evaluate beauty.

True ○ False ●

9 – Efficient techniques to measure the psychological profile of a user are making use of at least 25 questions.

True ○ False ● (10 are usually enough)

10 – Male/Female disambiguation is the first stage of any subjective profiling for aesthetic evaluation.

True ○ False ●