How Convolutional Neural Networks Deal with Aliasing

Antônio H. Ribeiro, Thomas B. Schön
Uppsala University, Sweden
{antonio.horta.ribeiro, thomas.schon}@it.uu.se

UPPSALA UNIVERSITET

Aliasing

Aliasing is a well-known side-effect of downsampling that may take place: it causes high-frequency components of the original signal to become indistinguishable from its low-frequency components.

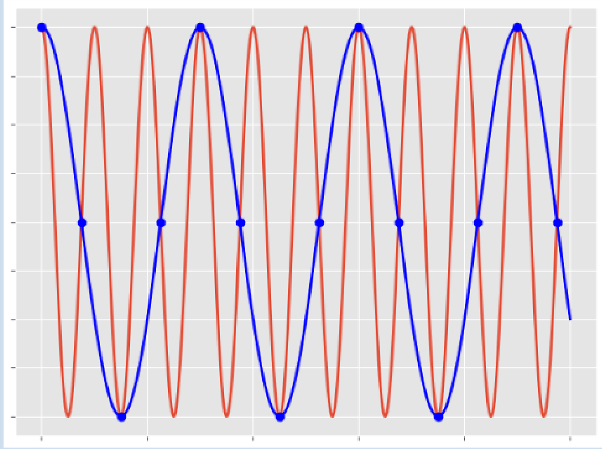

Fig.: Aliasing illustration

CNNs and downsampling

Standard convolutional architectures consist of stacked layers of operations that progressively downscale the image. While downsampling takes place in the max-pooling layers or in the strided-convolutions in these models, there is no explicit mechanism that prevents aliasing from taking place.

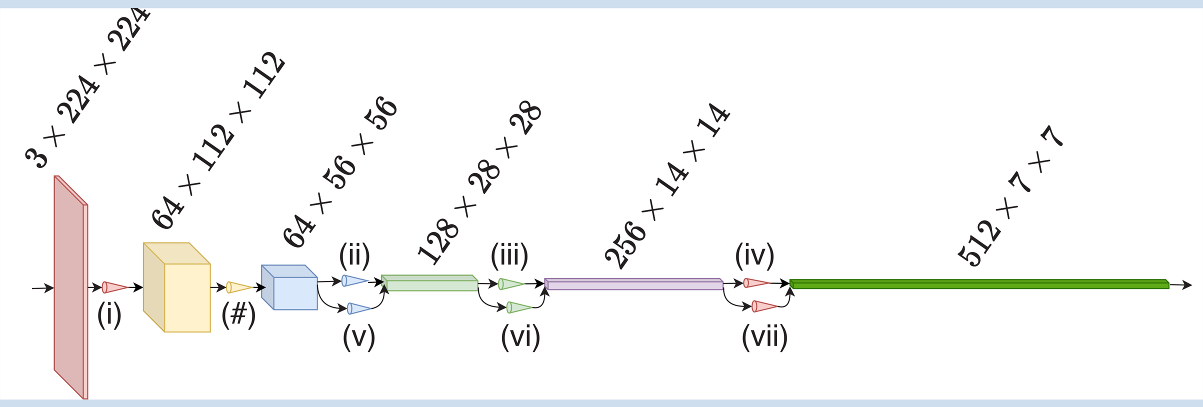

Fig.: Dimensions of tensors in a ResNet34.

Previous work: antialiased-CNNss

C. Vasconcelos, H. Larochelle, V. Dumoulin, *et al.*, "An Effective Anti-Aliasing Approach for Residual Networks," *arXiv:2011.10675*, Nov. 2020.

X. Zou, "Delving Deeper into Anti-aliasing in ConvNets," en, in *Proceedings of the 31st British Machine Vision Virtual Conference (BMVC)*, 2020.

R. Zhang, "Making Convolutional Networks Shift-Invariant Again," in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, arXiv: 1904.11486, Jun. 2019.

Our work takes a different route: rather than trying to explicitly include anti-aliasing mechanisms in a CNN, we aim to explain how these architectures still manage to be successful without it.

How do convolutional neural networks (CNNs) manage to be successful without explicit anti-aliasing mechanisms?

1. **Can CNNs resolve between oscillations at its input?** Aliasing results in loss of information and the inability to distinguish between frequency components. Does this happen in a CNN?
2. **Do CNNs learn anti-aliasing filters?** We try to assess if CNNs actually prevent aliasing from taking place in the intermediate layers. Convolutional layers are, in principle, capable of implementing anti-aliasing filters. Hence, even if these are not hard-coded in the structure, CNNs could learn it.

Can CNNs resolve between all oscillations?

Experiment 1. The model receive as input an oscillatory two-dimensional signal x and need to classify it according to the frequency (ω_1, ω_2). Here, each component ω_i is the frequency of the oscillations in one direction.

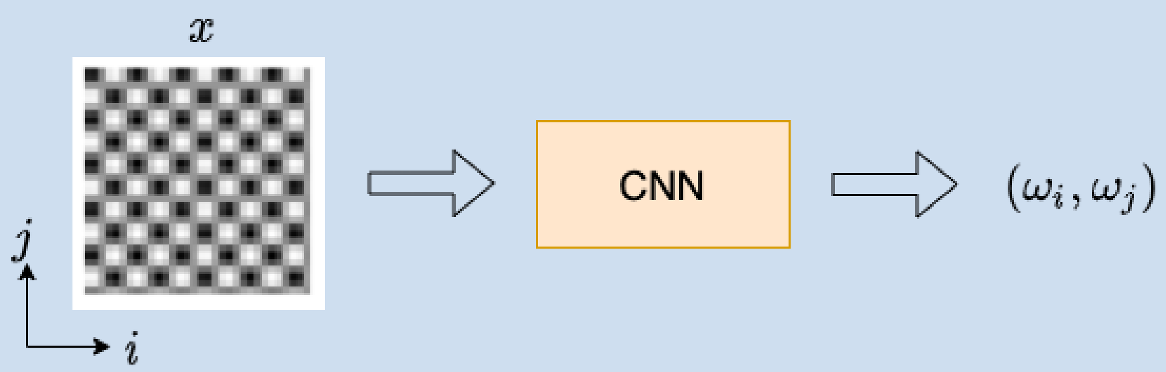

Fig.: Illustration of the classification problem.

The training and validation sets consist of pairs $\{x, \omega\}$, where ω contain 400 uniformly spaced frequencies.

What Experiment 1 tell us about aliasing? Every time the signal x is downsampled by a factor of two, the number of frequencies that can be distinguished drops by a factor of four due to aliasing. For instance, the oscillatory patterns below would become indistinguishable after being downsampled by a factor of 2.

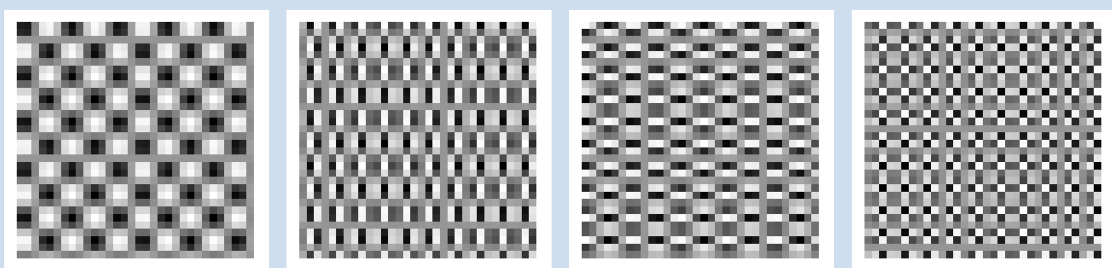

Fig.: Four oscillatory patterns. If the image is downsampled by 2 they become indistinguishable.

For the CNN model, the downsampling operations happen at intermediate layers and the goal of this task is to assess what is the resolution CNNs can resolve between distinct frequency components at the input.

The role of redundancy in distinguishing oscil.

Results (Experiment 1). The ResNet model performs better then fully connected networks, despite downsampling operations taking place. Increasing the number of channels can improve the performance. Increasing the depth, however, does not help much.

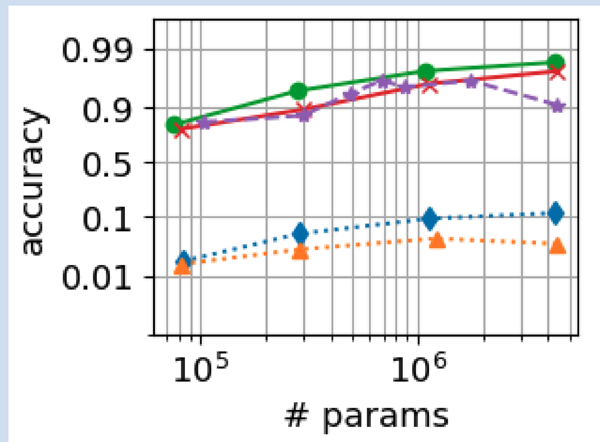

Fig.: Accuracy *vs* number of parameters.

Fully connected neural network with 1 hidden layer (blue dotted line) and 2 layers (orange dotted line) ploted together with Resnet with fixed depth and and increasing # of channels (purple dashed line); and with fixed # of channels and increasing depth channel: both for constant proportions (green line) and for increasing proportions (red line).

Quantifying aliasing

Let X' be the Discrete Fourier Transform (DFT) of the signal immediately after the downsampling operation. To each value of X' we attribute one of four mutually exclusive categories {no pass, non-aliased, aliased, aliased-tangled}

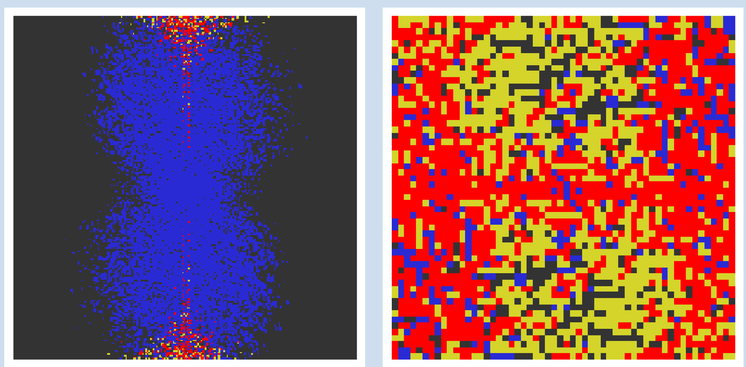

Fig. DFT of ResNet34 intermediate tensors classified as above (evaluated on ImageNet).

Do CNNs learn anti-aliasing filters?

Experiment 2. The figure below displays the fraction of DFT points in each category averaged over all channels and over all examples for ImageNet and for the task of Experiment 1. The outer chart shows the—equally weighted—average over the intermediate signals under consideration and the inner chart shows the individual contributions of each intermediate signal to the average. Aliasing occurs in, roughly, half of the frequency components.

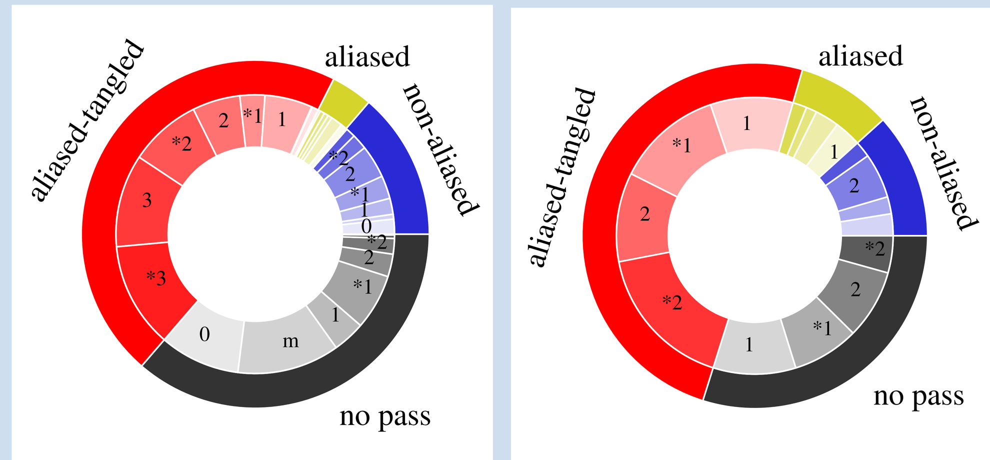

Fig. Fraction of samples suffering aliasing. On the left, for ImageNet; on the right, for the task of classifying the oscillations (Experiment 1).

Does it impact performance? One interesting question is whether examples that are incorrectly classified suffer aliasing to a different extent to those that are correctly classified. As shown in the figure below this does not seem to be the case.

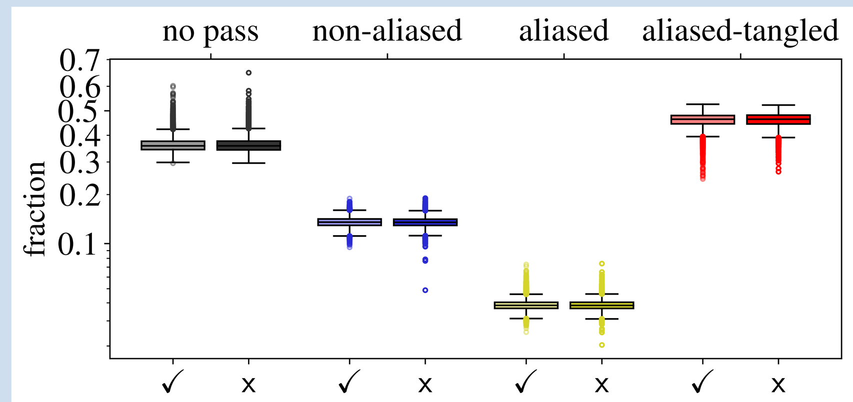

Fig. Correct (x) *vs* incorrect (✓) classified examples in ImageNet test set.

Conclusion

Our experiments indicate that **CNNs can resolve between oscillations at its input** but that they **don't learn anti-aliasing filters**. Nyquist's sampling theorem gives a *sufficient* criterion for guaranteeing donwsampling the signal without loosing information. Later developments have shown that it is possible to reconstruct signals downsampled below the Nyquist rate (i.e. compressive sensin). In the case of CNNs, the possibility of reconstruction is simplified by the channel redundancy.