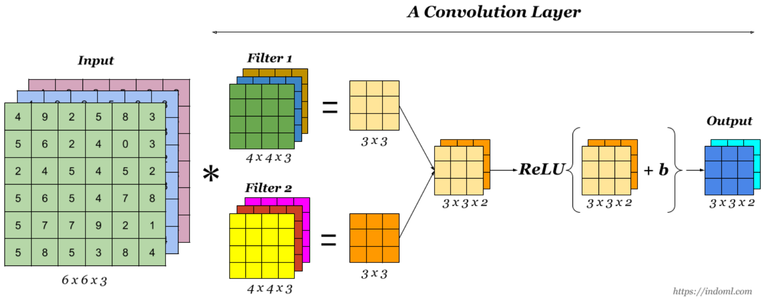
CERFACS

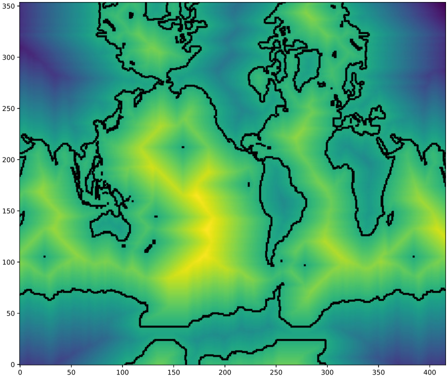
CENTRE EUROPÉEN DE RECHERCHE ET DE FORMATION AVANCÉE EN CALCUL SCIENTIFIQUE

Estimation of normalization coefficient with a CNN

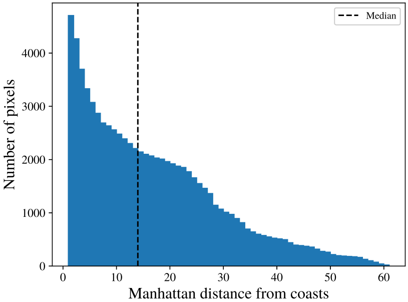
Folke, Guillaume



One convolution layer



Sign distance map



Distribution of ocean pixels by distance from the coast

- 1 input layer:
  - 3×3 convolution
  - 4 input channels, 64 output channels
  - Batch normalization
  - Activation function: Exponential Linear Unit (ELU)
- 8 hidden layers:
  - 3×3 convolution
  - 64 input channels, 64 output channels
  - Batch normalization
  - ELU
- 1 output layer
  - 3×3 convolution
  - 64 input channels, 1 output channel
- Skip connections

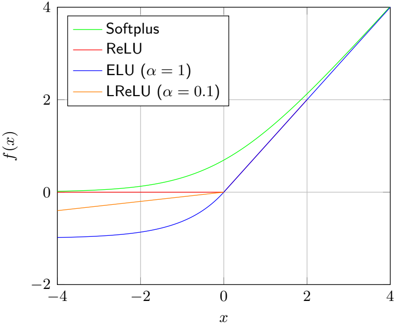
- Loss function: $\mathbb{E}[\tilde{\varepsilon}]$ with $\tilde{\varepsilon}$ defined component wise by $\tilde{\varepsilon}_i = (\frac{\hat{\gamma}_i^2 - \gamma_i^2}{\gamma_i^2})^2,\ \forall i \in$ Ocean cells
- 2.99×10^5 trainable parameters
- 5×10^4 epochs (number of passes on the entire training dataset)

$$\text{ReLU}(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{if } x \leq 0 \end{cases}$$

$$\text{LReLU}(x) = \begin{cases} x, & \text{if } x > 0 \\ \alpha x, & \text{if } x \leq 0 \end{cases}$$

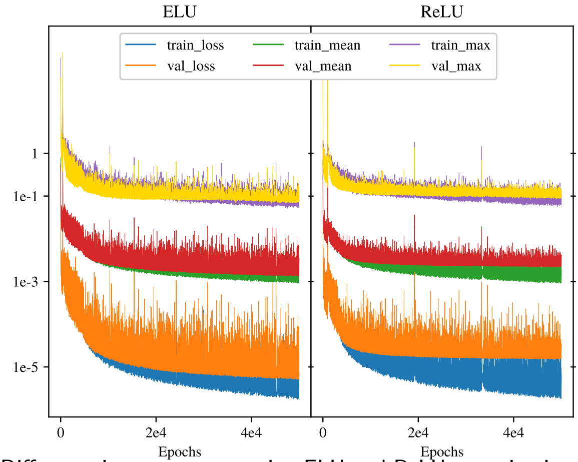
$$\text{ELU}(x) = \begin{cases} x, & \text{if } x > 0 \\ \alpha(e^x - 1), & \text{if } x \leq 0 \end{cases}$$

$$\text{Softplus}(x) = \log(1 + e^x)$$



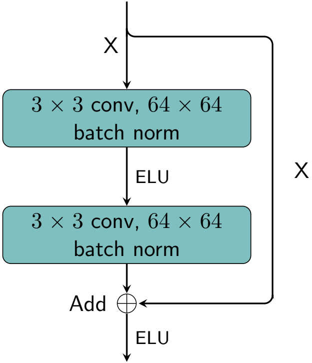
Different activation functions

- ELU and Softplus gave better results than ReLU
- The best results were obtained using ELU
- LReLU was worse

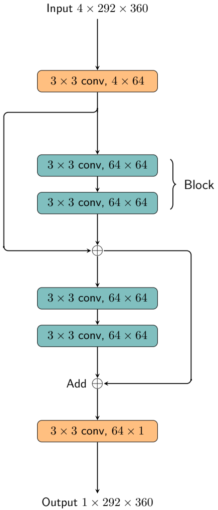


Difference in convergence using ELU and ReLU as activation functions

- 10 layers with the same input/output layers
- 8 hidden layers $\mapsto$ 4 blocks with 2 layers each
- Each block adds its input to its output

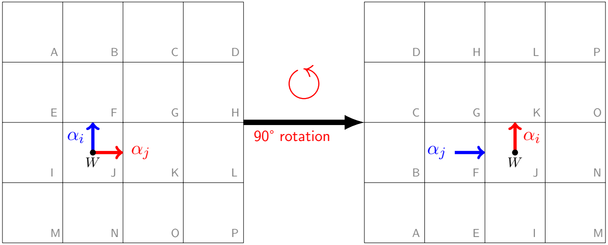


Block details



Architecture representation with 2 blocks and skip connections

- Data augmentation aims to artificially increase the amount of data by transforming existing data
- We tried **rotations** and **flips**. Further modifications are necessary for α_i and α_j



Representation of the 90° rotation

- But it did not work

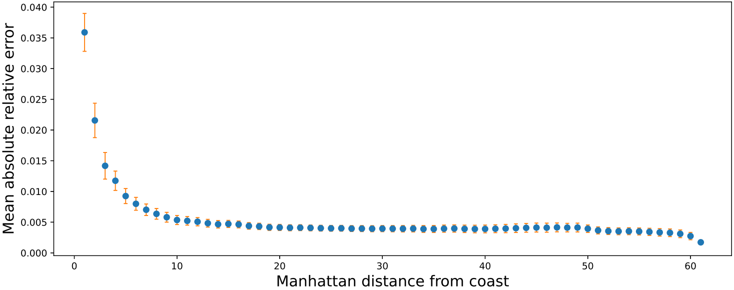
Randomized (10^4):

- For one 1 sample:
- Mean absolute relative error: 1.1 %
- Max absolute relative error: 5.6 %
- Quantile 99,99 of the relative error: 5.4 %

CNN:

- For one 90 samples:
- Mean absolute relative error: 0.65 %
- Max absolute relative error: 2.81 %
- Quantile 99,99 of the relative error: 1.90 %

- For one 190 samples:
- Mean absolute relative error: 0.38 %
- Max absolute relative error: 4.27 %
- Quantile 99,99 of the relative error: 1.87 %



Mean absolute relative error in relation to the distance to the coast

- On **kraken**: gpu partition (Tesla V100-PCIE-16GB)
- Training duration: 9 hours to 16 hours (2.99×10^5 to 7.45×10^5 trainable parameters). But it only has to be performed **once**.
- Computation speed on 1 sample: 1 to 2 seconds

- The CNN provides convincing results and is better than the randomization method in the context of the experiment
- We thought to perform the training on sections of the image
- Expand to the 3-dimensional case

Thank you for your attention !