

Week 6: take-home messages

- CNN architecture: define it according to the dataset size, and the complexity and nature of the task.
- Optimiser: `torch.optim.Adam` (fast convergence) vs `torch.optim.SGD` (better generalisation).
- Criterion: will depend on the task, e.g., MSE for regression and cross-entropy for classification.
- Validation set to terminate the training with early stopping.
- One can access the learnt filters and recover patterns in the data (at different scales!).
- Pooling reduces the number of parameters and thus addresses overfitting.
- When dealing with images, smaller kernel sizes (3×3 to 5×5) are a popular choice over larger sizes.