UNIVERSITY OF CAMBRIDGE

Automatic inference of cross-modal connection topologies for X-CNNs

Laurynas Karazija, laurynas.karazija@cantab.net

Computer Laboratory, University of Cambridge, UK

International Symposium on Neural Networks 27 June 2018

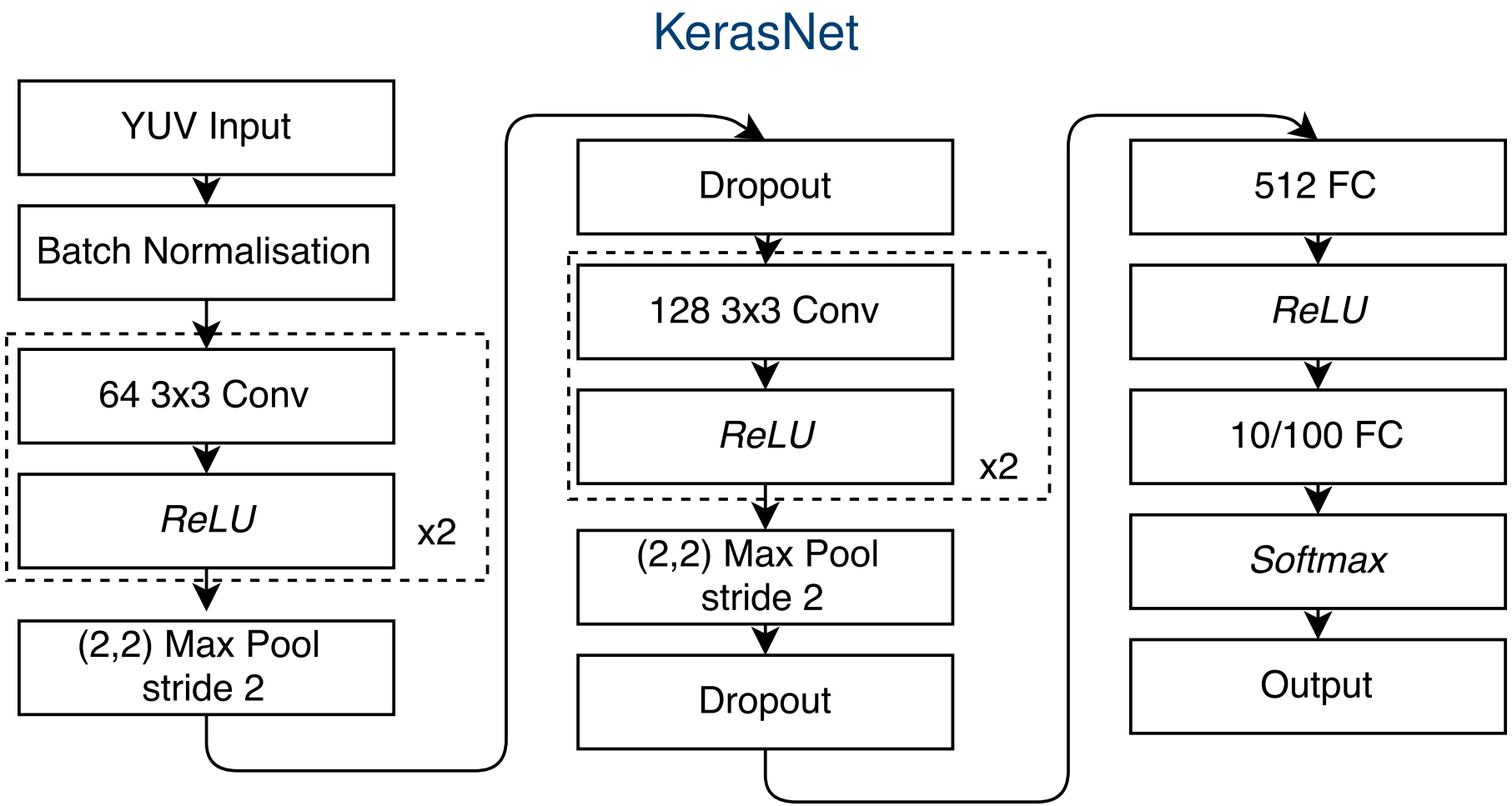
Overview

- Overview of the problem
- Contributions
- Cross-modal architectures
- Cross-modal connection
- Connection *weight*
- Xsertion
- Results
- Conclusions

The problem

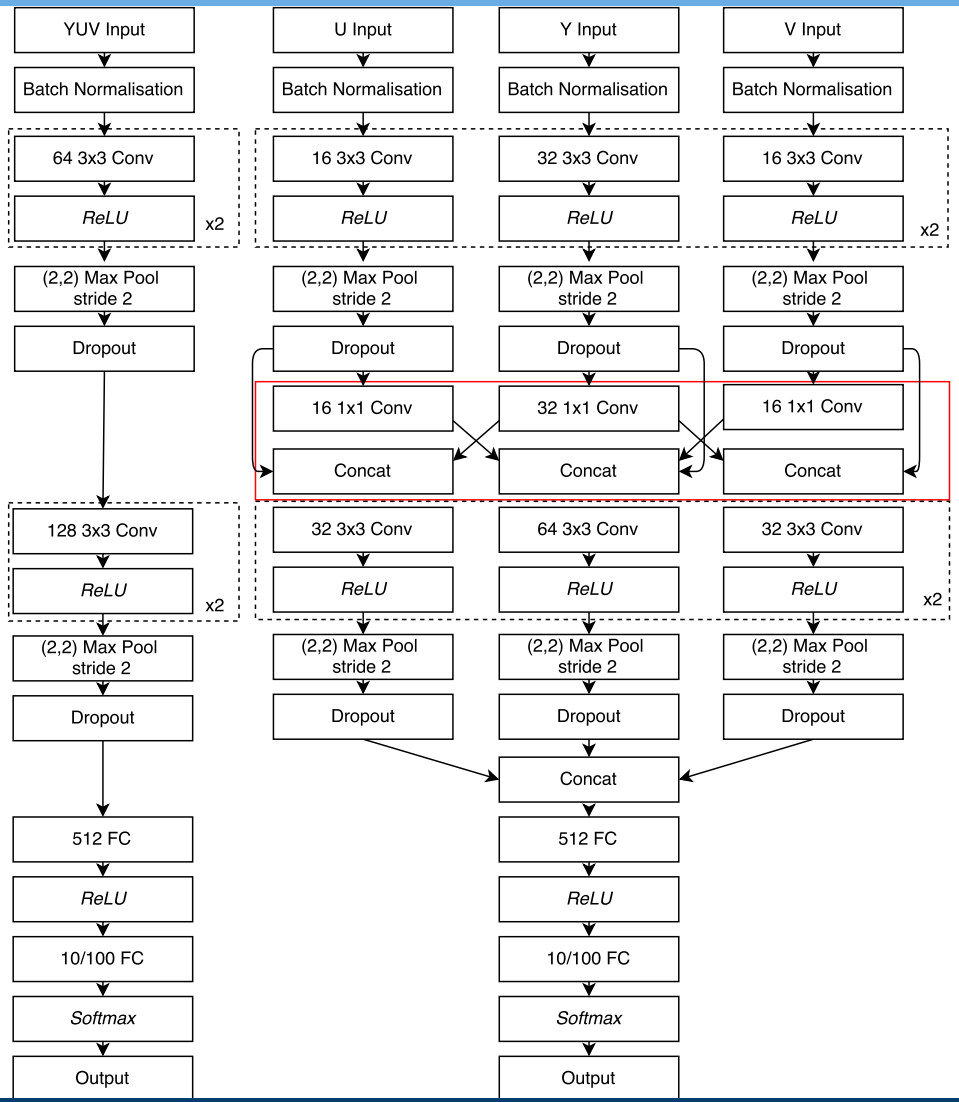
- Neural networks are great but require a lot of examples.
- Many field have very few examples but the data is wide – multiple *modalities* are present.
- X-CNNs provide means to address this but are difficult to design.

The problem: CNN



The problem: CNN → X-CNN

KerasNet



XKerasNet

Contributions

- Investigation into cross-modal architecture.
- Experimentation with X-CNN structure.
- Explanation of various parts of topology and their impact.
- Introduction of heuristics to decide topology.
- Proposal of a combined learning procedure to build the networks automatically.

Cross-modal connections

are a crucial aspects of the architecture. They join two super-layers together forming a connection from the *origin* super-layer to the *destination* super-layer.

- Provide extra-modal context to the feature detector aimed at a single modality via *feature transfer*.
- The key aspect of cross-modal networks that enables them to work in low data-availability environments.

Cross-modal connections

Cross-modal connections apply 1x1 convolution to provide additional information from other modalities. On lower lever, they:

- Apply an affine transformation of features,
- Compress the information transferred along the connections,
- Provide gating during training.

Formalising connections

Connection weight

- Let l_a, l_b be super-layers for modalities A and B.
- Connection weight is such a number $w_{l_a,l_b} \in [0, 1]$, so that $w_{l_a,l_b} \geq 0.5$ if modality A is more informative than modality B.

Formalising connections

Connection weight

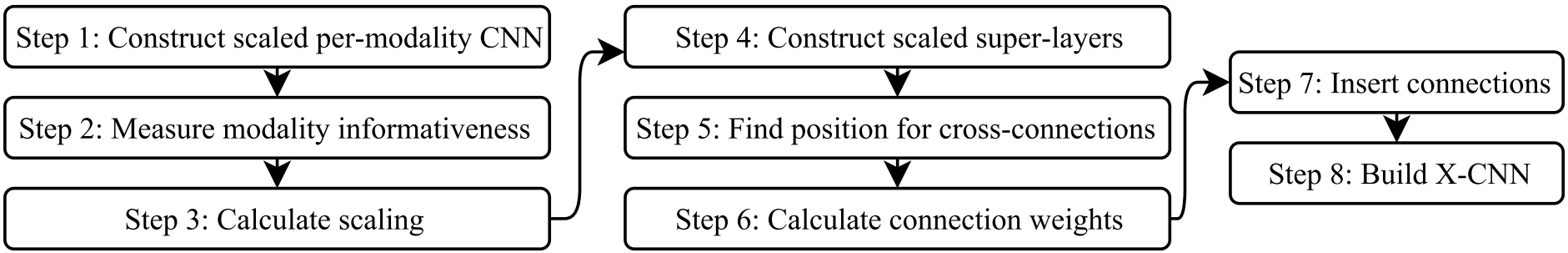
- Let l_a, l_b be super-layers for modalities A and B.
- Connection weight is such a number $w_{l_a,l_b} \in [0, 1]$, so that $w_{l_a,l_b} \geq 0.5$ if modality A is more informative than modality B.

Formulation

- Let n_{l_a}, n_{l_a} be some measures of *informativeness*.

$$w_{l_a,l_b} = \frac{n_{l_a}^{\beta}}{n_{l_a}^{\beta} + n_{l_b}^{\beta}} \tag{1}$$

Xsertion


Step 1: Construct scaled per-modality CNN
Step 2: Measure modality informativeness
Step 3: Calculate scaling
Step 4: Construct scaled super-layers
Step 5: Find position for cross-connections
Step 6: Calculate connection weights
Step 7: Insert connections
Step 8: Build X-CNN

Results

CIFAR-10

Model \ $p\%$	20% (%)	40% (%)	60% (%)	80% (%)	100% (%)
FitNet	75.47 ± 0.32	82.02 ± 0.18	84.98 ± 0.20	86.22 ± 0.19	87.42 ± 0.05
XFitNet	76.56 ± 0.24	82.43 ± 0.07	85.11 ± 0.19	86.23 ± 0.18	87.42 ± 0.08
Xsertion	$\mathbf{77.35 \pm 0.15}$	$\mathbf{82.66 \pm 0.09}$	$\mathbf{85.43 \pm 0.12}$	$\mathbf{86.78 \pm 0.16}$	$\mathbf{87.77 \pm 0.22}$

CIFAR-100

Model \ $p\%$	20% (%)	40% (%)	60% (%)	80% (%)	100% (%)
FitNet	29.29 ± 1.69	40.91 ± 2.48	50.94 ± 0.51	55.47 ± 0.96	58.92 ± 0.60
XFitNet	36.17 ± 0.27	48.02 ± 0.72	54.18 ± 0.36	57.98 ± 0.33	60.32 ± 0.29
Xsertion	$\mathbf{38.59 \pm 0.37}$	$\mathbf{50.11 \pm 0.30}$	$\mathbf{55.48 \pm 0.41}$	$\mathbf{59.06 \pm 0.63}$	$\mathbf{61.67 \pm 0.31}$

Results: Residual Learning

- Applied to a variant of residual in residual network.
- Contained 12 residual blocks and used preactivations.
- Xsertion produced improvement 85.72% $\rightarrow$ 88.81% and 55.43% $\rightarrow$ 61.33% on CIFAR-10/100 respectively.

Learn connections and parameters simultaneously

Perform *gradient descent* in a parameter space that includes *all* potential connections. Use another procedure, gradient ascent, to *restrict* and *optimise* a set of axes for the gradient descent. This way both X-CNN the connections and parameters are trained.

Summary

- A method to automatically infer and construct cross-modal convolutional neural networks was produced.
- The models perform better than hand-constructed ones, taking less time to build.
- The library provides experimentation platform for ideas in cross-modality, whilst also impelling a way to apply a bleeding-edge idea in deep learning, inviting similar approaches to be taken in other research.

Thank you! Questions?

Laurynas Karazija
laurynas.karazija@cantab.net