UNIVERSITY OF CAMBRIDGE

X-CNN: Cross-modal convolutional neural networks for sparse datasets

Petar Veličković[1,3], Duo Wang[1], Nicholas D. Lane[2,3] and Pietro Liò[1]

[1]Computer Laboratory, University of Cambridge, UK
[2]Department of Computer Science, University College London, UK
[3]Nokia Bell Labs, UK

IEEE Symposium Series on Computational Intelligence 2016 (CIEL) 7 December 2016

Introduction

Sparse data

- Deep neural networks have become a *major success story* of AI—primarily on problems involving “big data”.
- Properly applying them to small data environments quickly becomes difficult given the large number of parameters.
- In this talk I will present our method for exploiting data *width*—data modalities within an otherwise small dataset.
- Think clinical studies: few patients (rows), rich patient history (columns).

Related work

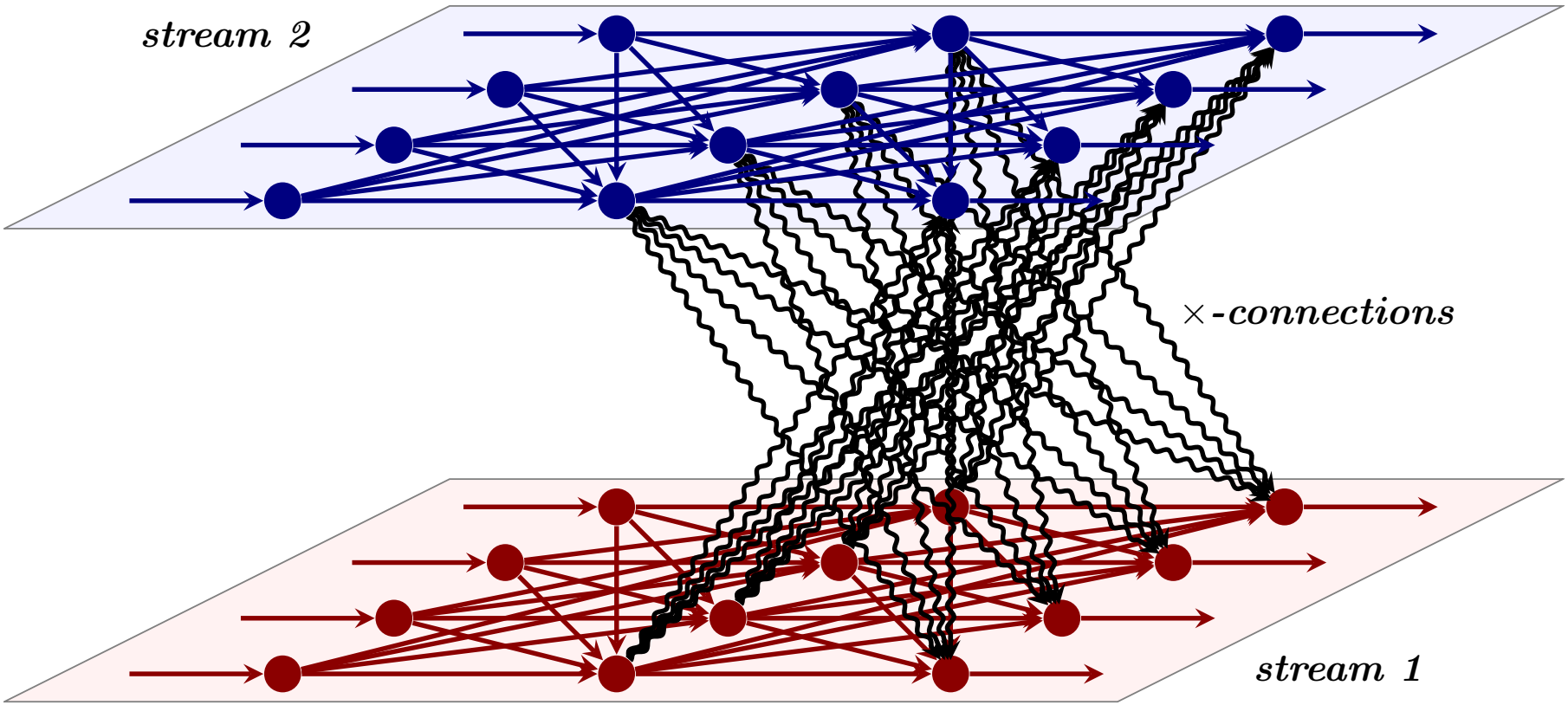
- We emphasise two previous attempts to learn from multiple data modalities (typically A/V or image/tags):
  - Ngiam *et al.* (2011), ICML (using autoencoders)
  - Srivastava and Salakhutdinov (2012), NIPS (using DBMs)

- The models presented employ separate processing streams for each modality, culminating with joint representation layer(s).

Cross-connections

- We generalise further by allowing further *cross-connections* anywhere within the individual modal processing streams.
- Now the layers specialised on a particular modality can periodically *exchange information*.
- Effectively preserves the benefits of a fully unrestricted network while still providing a decrease in parameters!

Cross-connection example



Cross-modal networks

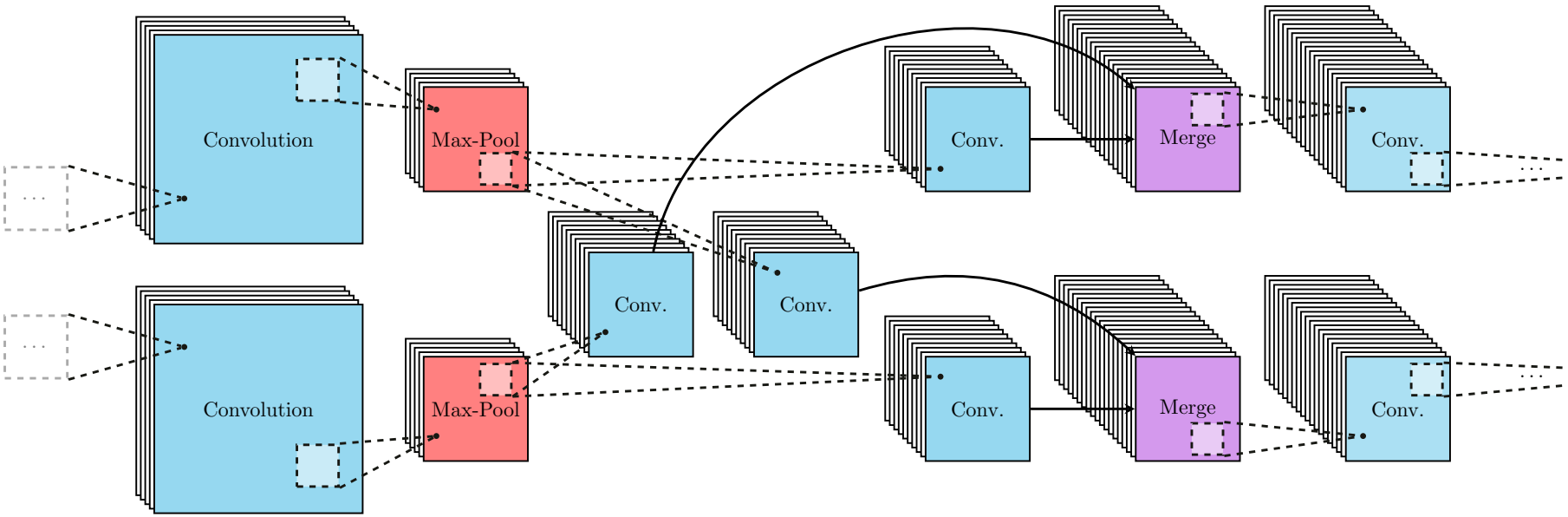
- We use the term **cross-modal neural network** to describe any neural network with distinct *subnetworks* processing separate (not necessariy mutually disjoint) modalities of the input data, while allowing for *cross-connections*.
- As with most proposed neural network architectures, this construction is *biologically inspired*.
- Evidence of cross-modality between the human auditory and visual cortices has been published on several occasions:
  - Eckert *et al.* (2008), Human Brain Mapping;
  - Beer *et al.* (2011), Experimental Brain Research;
  - Yang *et al.* (2015), PLoS ONE.

Results

X-CNNs

- We applied the cross-modal construction to convolutional neural networks—giving rise to *cross-modal convolutional neural networks* or **X-CNNs** for short.
- We considered the popular task of *image classification*, treating each view into an image (e.g. RGB/YUV channels) as a separate modality.
- We apply cross-connections after each *pooling* (downsampling) layer, for a typical four-layer CNN and a sophisticated FitNet4 (Romero *et al.* (2015), ICLR)

X-CNN cross-connections

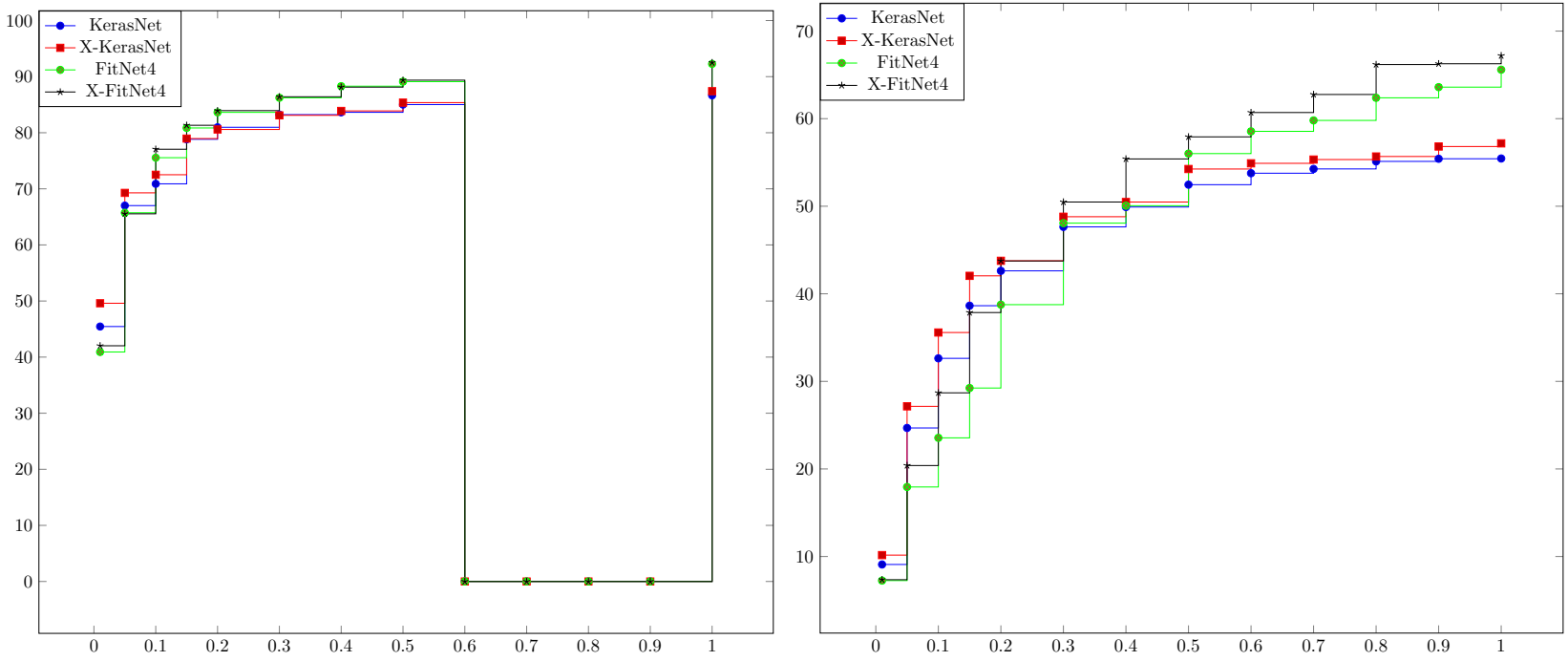


The cross-connections employ 1×1 convolutions, corresponding to *exchanging linear combinations of the feature maps*.

Evaluation of X-CNN

- We evaluate the constructed X-CNNs on the CIFAR-10 and CIFAR-100 benchmarks (transformed into YUV).
- We compare the X-CNNs' performance against their baseline single-stream CNNs (adjusting layer sizes to make sure that they have comparable numbers of parameters).
- For evaluating the sparse data application, we train using only a subset of p% of the training data.
- We vary p in increments of 10%. For better coverage of the small data range, we also train on 1%, 5% and 15% of the data.
- Expectation: Higher performance of the X-CNN up to a particular data availability threshold.

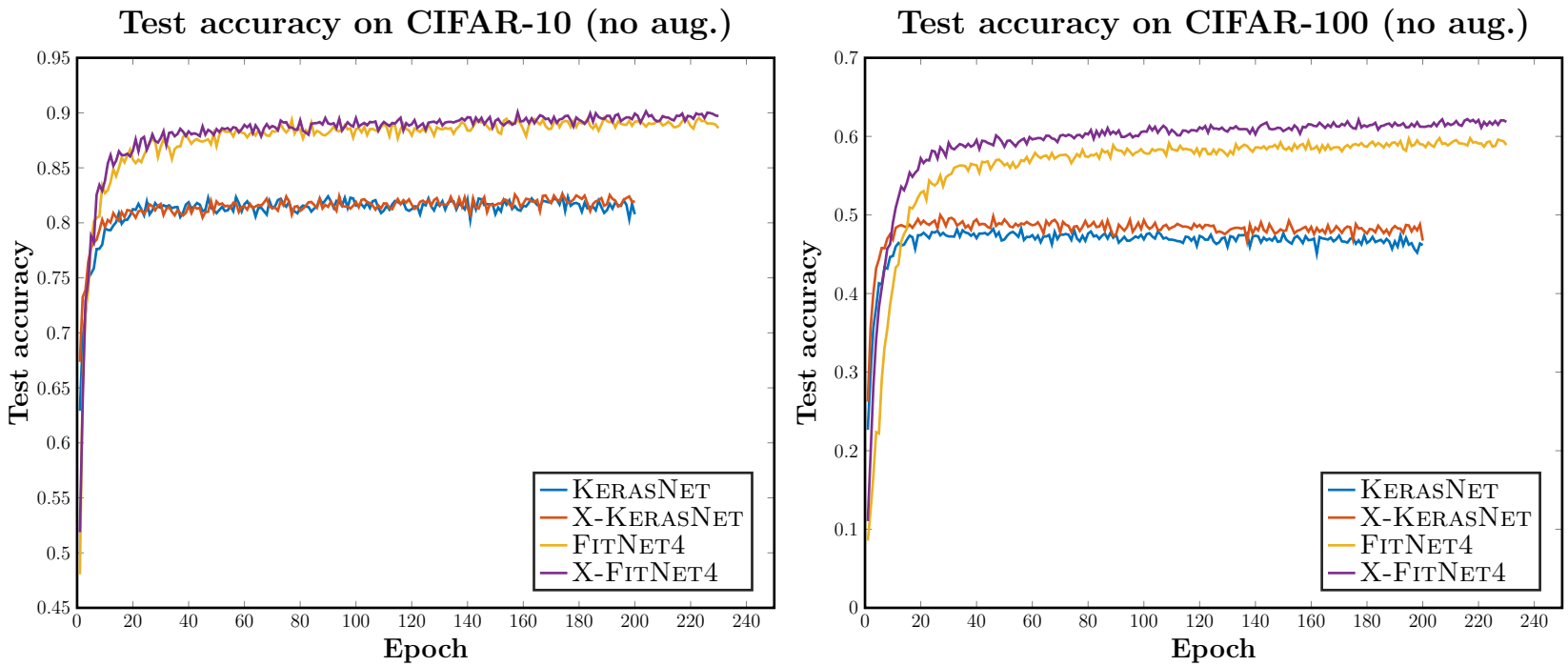
Test accuracy comparison



(a) CIFAR-10 (b) CIFAR-100

Figure: Plots of the test accuracy of the four CNN models against the percentage of the dataset used in training.

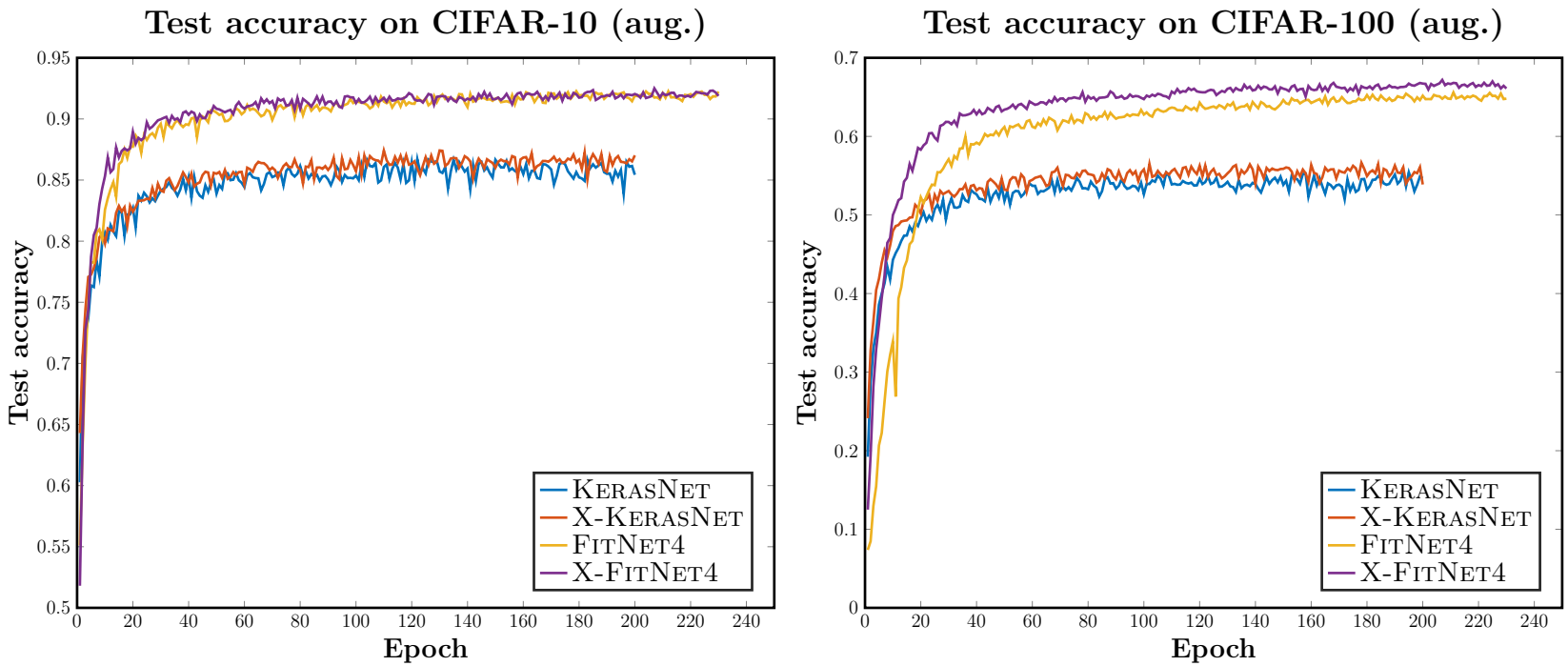
Training progress comparison



(a) CIFAR-10 (b) CIFAR-100

Figure: Plots of the test accuracy of the four CNN models against the number of training epochs (without data augmentation).

Training progress comparison



(a) CIFAR-10 (b) CIFAR-100

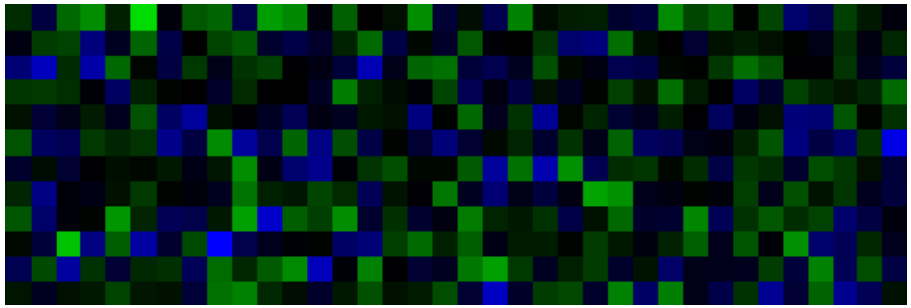
Figure: Plots of the test accuracy of the four CNN models against the number of training epochs (with data augmentation).

Results summary

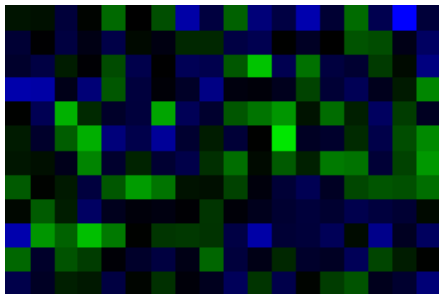
- These results demonstrate that X-CNNs *significantly outperform* their unrestricted variants on small datasets ($\sim$ 25 images per class), while remaining competitive (and often *better*) on large datasets ($\sim$ 5000 images per class).
  - We confirmed *statistical result significance* ($p < 0.05$) for data set sizes up to 15%, after retraining each model five times.
- It should therefore be a good idea to attempt a X-CNN variant on any such problem *regardless of whether big data is present*.
  - The effects of the method also compound well with *data augmentation* (where applicable).

Visualising cross-connections

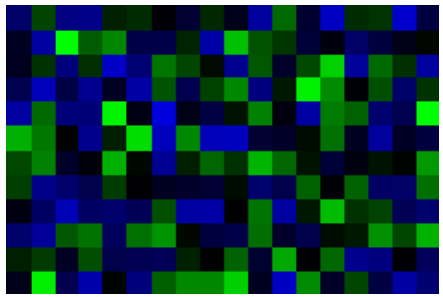
- Cross-connection weights are essentially 2-dimensional mapping matrices.
- We visualise by using colours proportional to weight values:
  - Green: positive weights
  - Blue: negative weights
- $\times$-connections selectively filter and combine features.



(a) Y $\rightsquigarrow$ U/V



(b) U $\rightsquigarrow$ Y

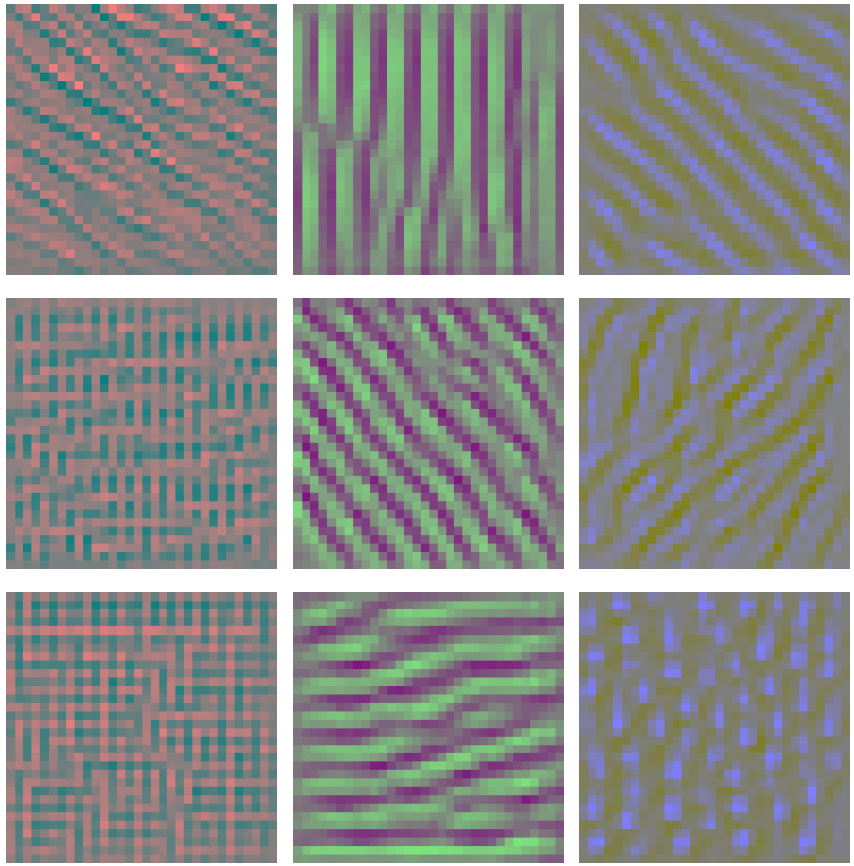


(c) V $\rightsquigarrow$ Y

Figure: Weight visualisation

Visualising cross-connections

- We visualise features that maximise activation of a particular $\times$-connection neuron (via gradient ascent on a white noise image).
- Cross-connections learn to combine low-level features (e.g. horizontal/vertical).
- Y stream features have higher frequency than U/V ($\sim$*mimicking human vision*).



(a) Y$\rightsquigarrow$U/V (b) U$\rightsquigarrow$Y (c) V$\rightsquigarrow$Y

Application: Clinical decision support system

Clinical decision support system: Components

- *X-Ray Computed Tomography* (**CT**)
  - Highly accurate anatomical localisation;
- *Magnetic Resonance Imaging* (**MRI**)
  - Good at distinguishing soft tissues;
- *Positron Emitting Tomography* (**PET**)
  - Highlight metabolic activity.

Application: Clinical decision support system

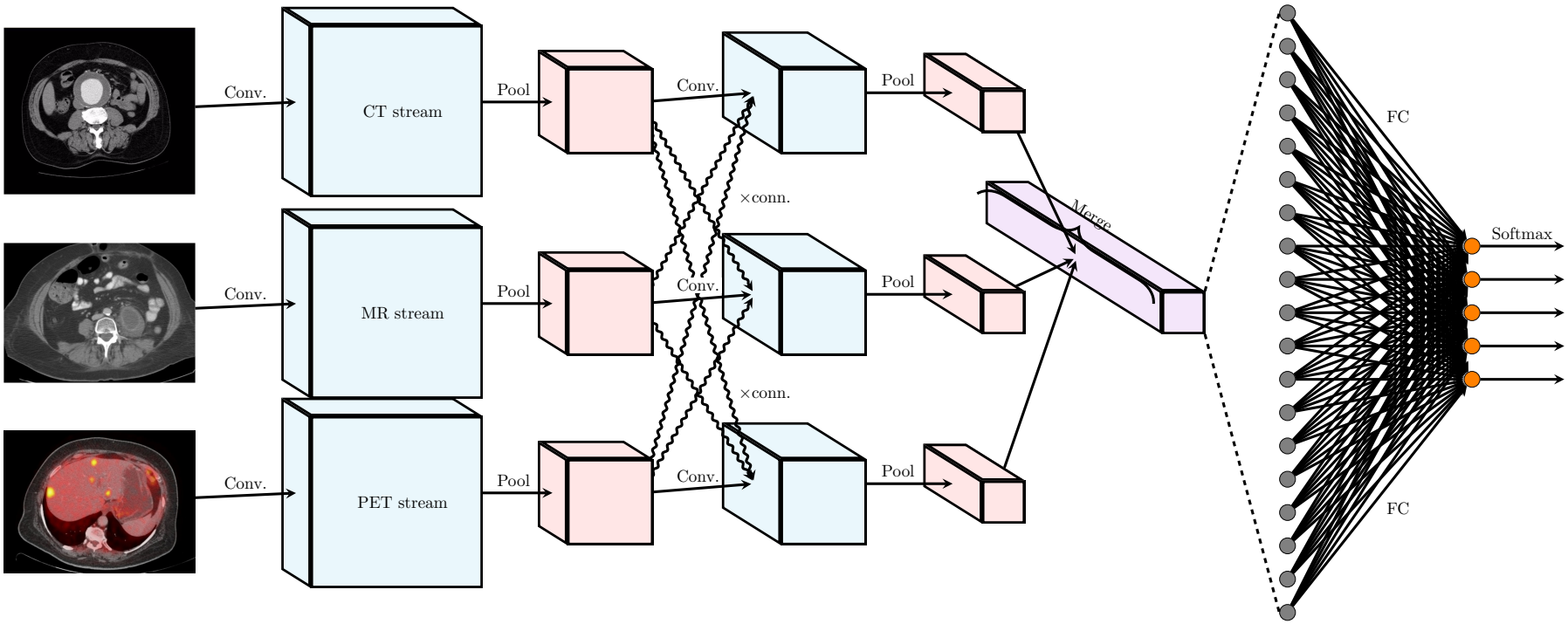


Figure: X-CNN applied to imaging-based tumour detection. Three streams of X-CNN consume CT, MRI and PET images as input.

Application: Clinical decision support system

- X-CNNs can produce enhanced results based on multimodal image analysis, given the inherent sparsity of the problem.
- Other patient information, including patient demographic data, medical history, genetic information, co-morbidity etc. can also be analysed within the model.
- A *clinical decision support system* can be developed to facilitate disease diagnosis and treatment by analysing all of the information available to clinicians.

Thank you!

Questions?

petar.velickovic@cl.cam.ac.uk

http://www.cl.cam.ac.uk/∼pv273/

https://github.com/PetarV-/X-CNN