embedded VISION SUMMIT 2018

Hybrid Semi-Parallel Deep Neural Networks (SPDNN) – example methodologies & use cases

XPERI

Peter Corcoran & Shabab Bazrafkkan

May 23, 2018

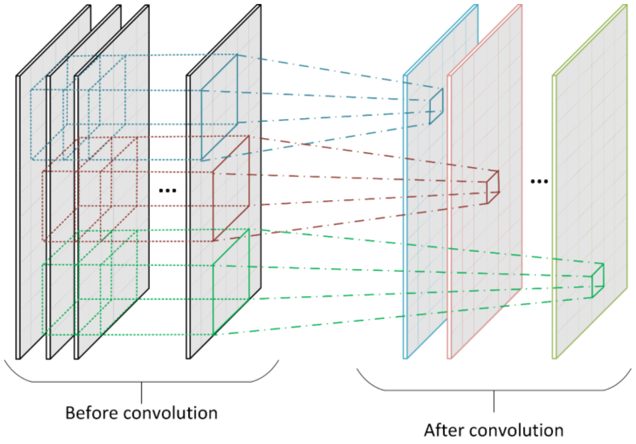
Overview of The Presentation

- **Background & Introduction**
- The Dataset Conundrum (*the back-story*)
  - Case Study – Facial Expression Datasets
- Hybrid Semi-Parallel Deep Neural Networks (SPDNN)
  - Merging the Parallel Networks
  - Some Case Studies
    - Case Study #1: Iris Segmentation
    - Case Study #2: MonoVision Depth-Map

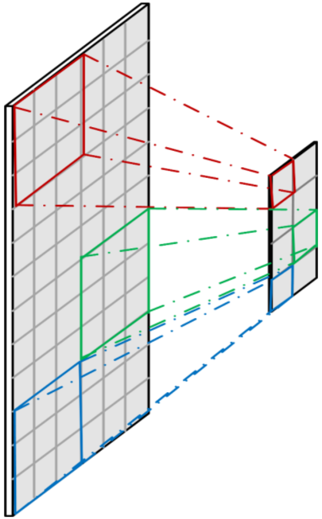
Background & Introduction

Conventional Deep Neural Networks #1

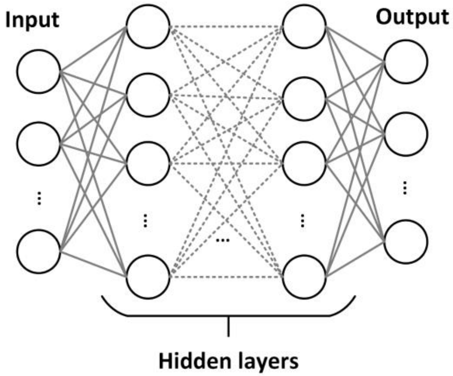
- Typical Deep NNs are comprised of Convolutional, Pooling, and Fully connected layers:



Convolutional Layers



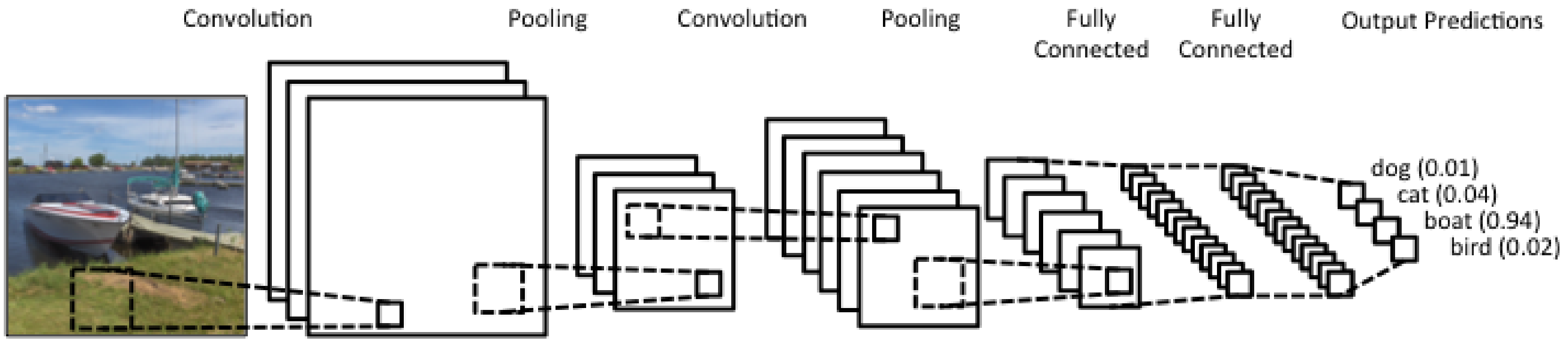
Pooling Layers



Fully Connected Layers

Conventional Deep Neural Networks #2

- And, they are ‘Deep’ … the network decision-making and discriminating power is typically improved by adding *more layers* to increase the depth …



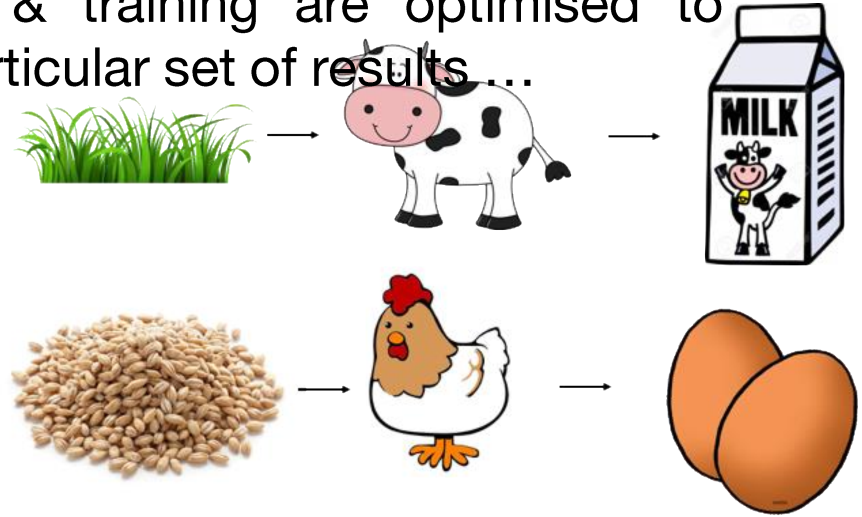
Overview of The Presentation

- Background & Introduction
- **The Dataset Conundrum (*the back-story*)**
  - **Case Study – Facial Expression Datasets**
- Hybrid Semi-Parallel Deep Neural Networks (SPDNN)
  - Merging the Parallel Networks
  - Some Case Studies
    - Case Study #1: Iris Segmentation
    - Case Study #2: MonoVision Depth-Map

The Dataset Conundrum

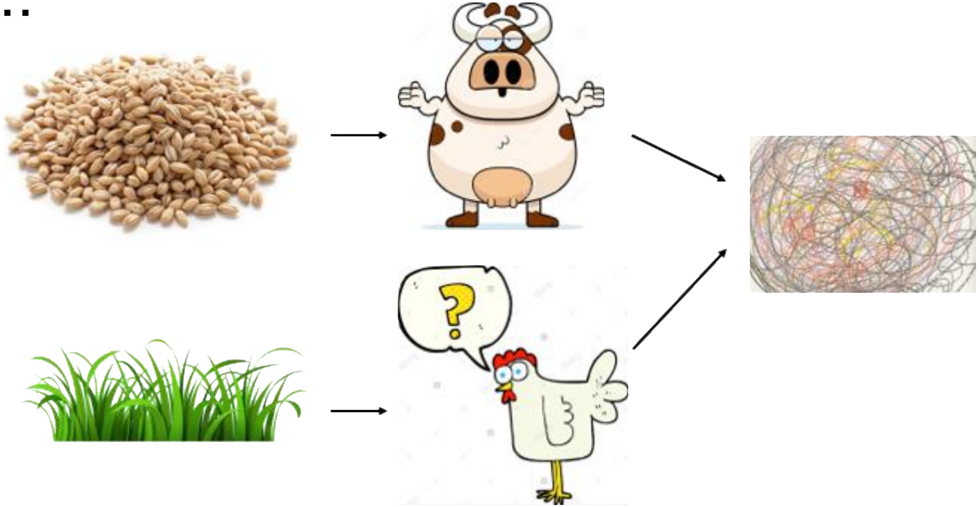
Networks & Datasets #1

- Now as researchers, we train our networks on a particular dataset …
- … and the network design & training are optimised to process that data to yield a particular set of results …



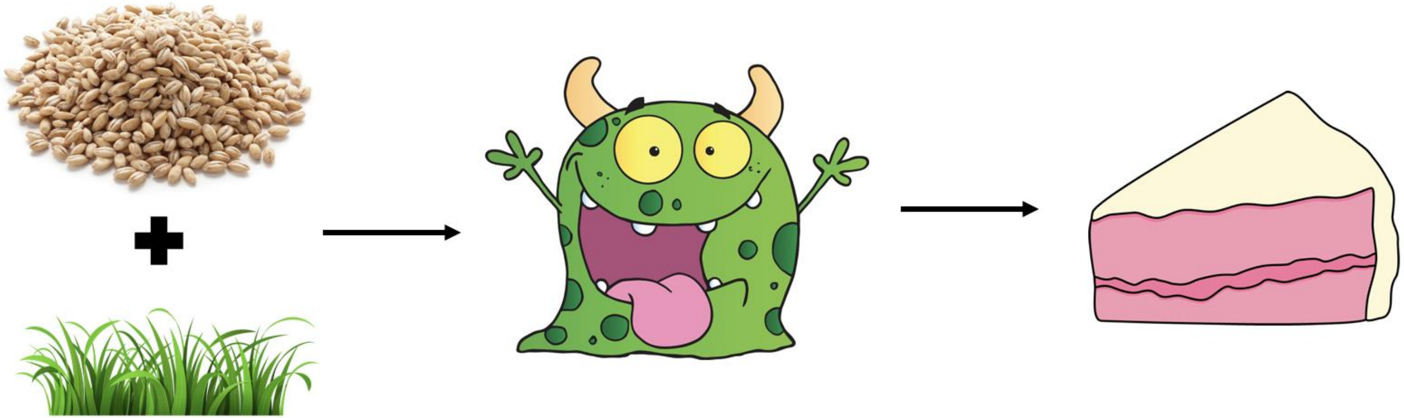
Networks & Datasets #2

- So what happens if you apply a deep neural network to a dataset that is significantly different from the one is was trained on …
- The result is not very good …



Networks & Datasets #2

- This line of thinking led us to experiment with some “similar, but different” datasets to see if we could create a ‘hybrid’ that could handle “similar, but different” datasets in a sensible way …



Case Study: Facial Expression Datasets

Facial Expression Datasets (samples # in dataset)

Database\Dataset	Train	Validation	Test
RaFD	13160	312	146
CK+	14392	304	187
JAFFE	1960	42	22

Network Designs for each individual Dataset

Layer	Kernel/Units	Size/Dropout Probability
Convolutional	16	3x3
Maxpool	N/A	2x2
Convolutional	8	3x3
Maxpool	N/A	2x2
Convolutional	8	3x3
Maxpool	N/A	2x2
Fully Connected	15	Dropout p=0.8
Fully Connected	7	Dropout p =0.5

Network designed for RaFD

Layer	Kernel/Units	Size/Dropout Probability
Convolutional	8	3x3
Maxpool	N/A	2x2
Convolutional	8	3x3
Maxpool	N/A	2x2
Convolutional	8	3x3
Maxpool	N/A	2x2
Fully Connected	7	Dropout p =0.5

Network designed for CK+

The JAFFE dataset is too small to train a separate DNN network for it. It was used only as a test/validation dataset.

Network Design for the Mixed Dataset

Layer	Kernel/Units	Size/Dropout Probability
Convolutional	16	3x3
Maxpool	N/A	2x2
Convolutional	13	3x3
Maxpool	N/A	2x2
Convolutional	10	3x3
Maxpool	N/A	2x2
Fully Connected	7	Dropout p =0.5

Network designed for the mixed dataset (RaFD , CK+ , JAFFE)

Test Results on these Networks

Network\error	Error for RaFD	Error for CK+	Error for JAFFE
Network 1 (RaFD)	6.84%	59.59%	72.73%
Network 2 (CK+)	21.23%	19.59%	50%
No JAFFE Net			NA
Network 3 (ALL DBs)	**4.1%**	**16.04%**	**13.36%**

S. Bazrafkan, T. Nedelcu, P. Filipczuk, and P. Corcoran, **"Deep Learning for Facial Expression Recognition: A step closer to a SmartPhone that Knows your Moods"**, in IEEE International Conference on Consumer Electronics (ICCE), 2017.

What did we learn?

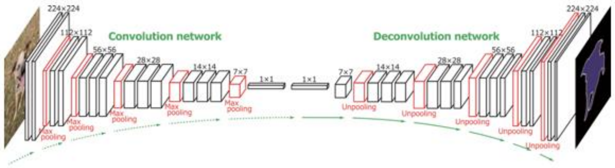
- The network for ‘combined dataset’ was retrained on a mixed set of samples …
  - This was a ‘compromise’ between the two networks optimised for the RaFD and CK+ datasets …
  - We hoped it would have similar performance on each network - i.e. it would have performance that was almost as good as the networks trained on each individual dataset … we did not expect it to perform better than the ‘dataset optimised networks
- In fact it <u>‘learned’ a better result</u> than the networks optimised for individual networks; so information from the ‘different’ datasets was able to improve the overall performance …

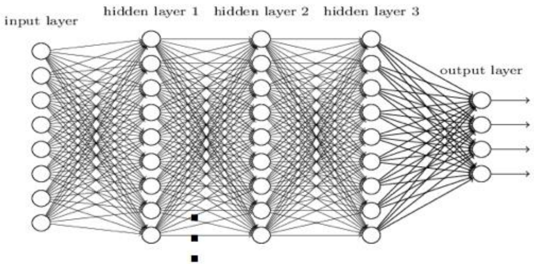
Overview of The Presentation

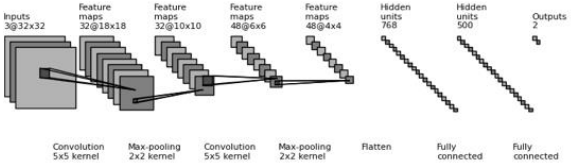
- Background & Introduction
- The Dataset Conundrum (*the back-story*)
  - Case Study – Facial Expression Datasets
- **Hybrid Semi-Parallel Deep Neural Networks (SPDNN)**
  - **Merging the Parallel Networks**
  - Some Case Studies
    - Case Study #1: Iris Segmentation
    - Case Study #2: MonoVision Depth-Map

Hybrid Semi-Parallel Deep Networks

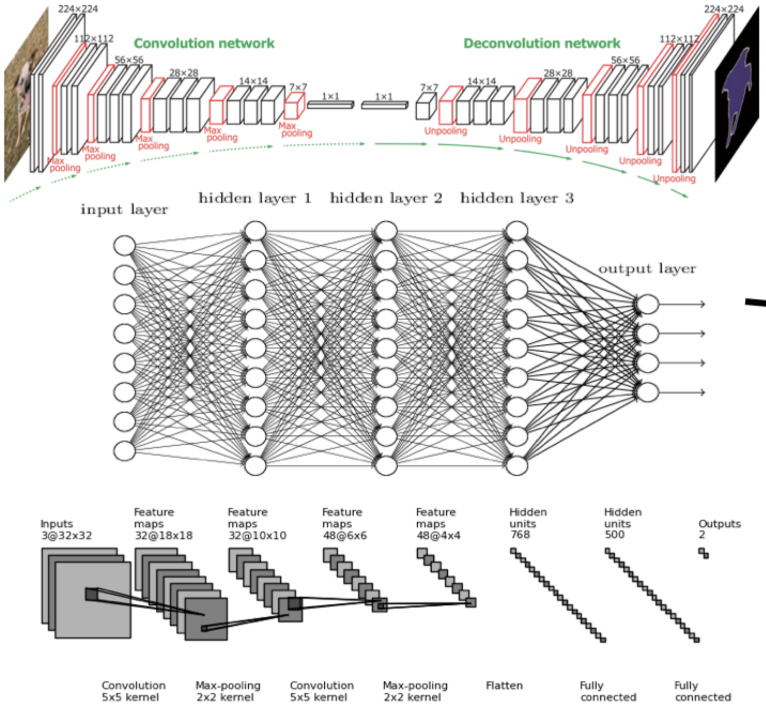
When there are many solutions in the Literature …







Can we combine these to obtain an improved solution?



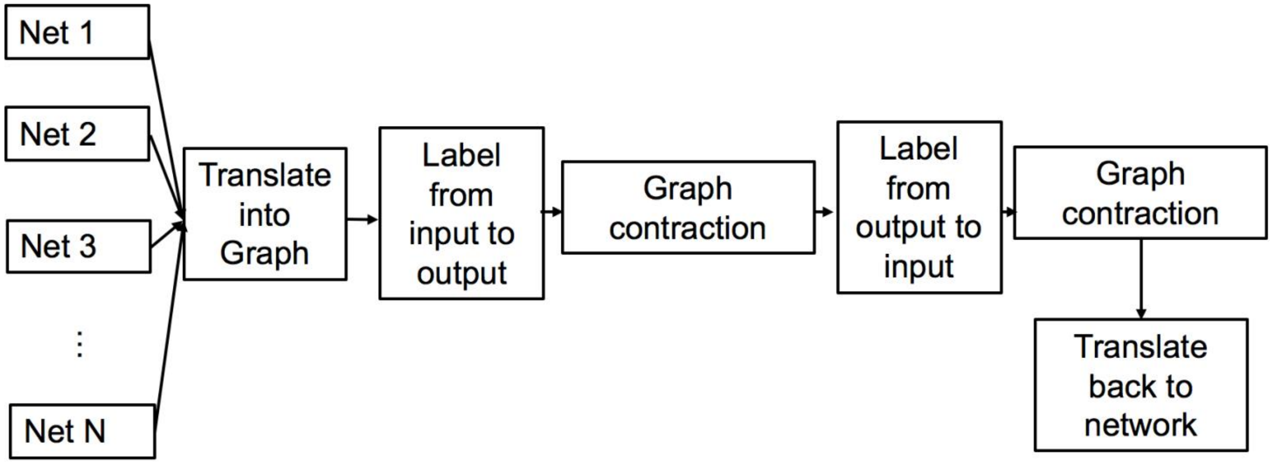
SPDNN

Hybrid Merged Network

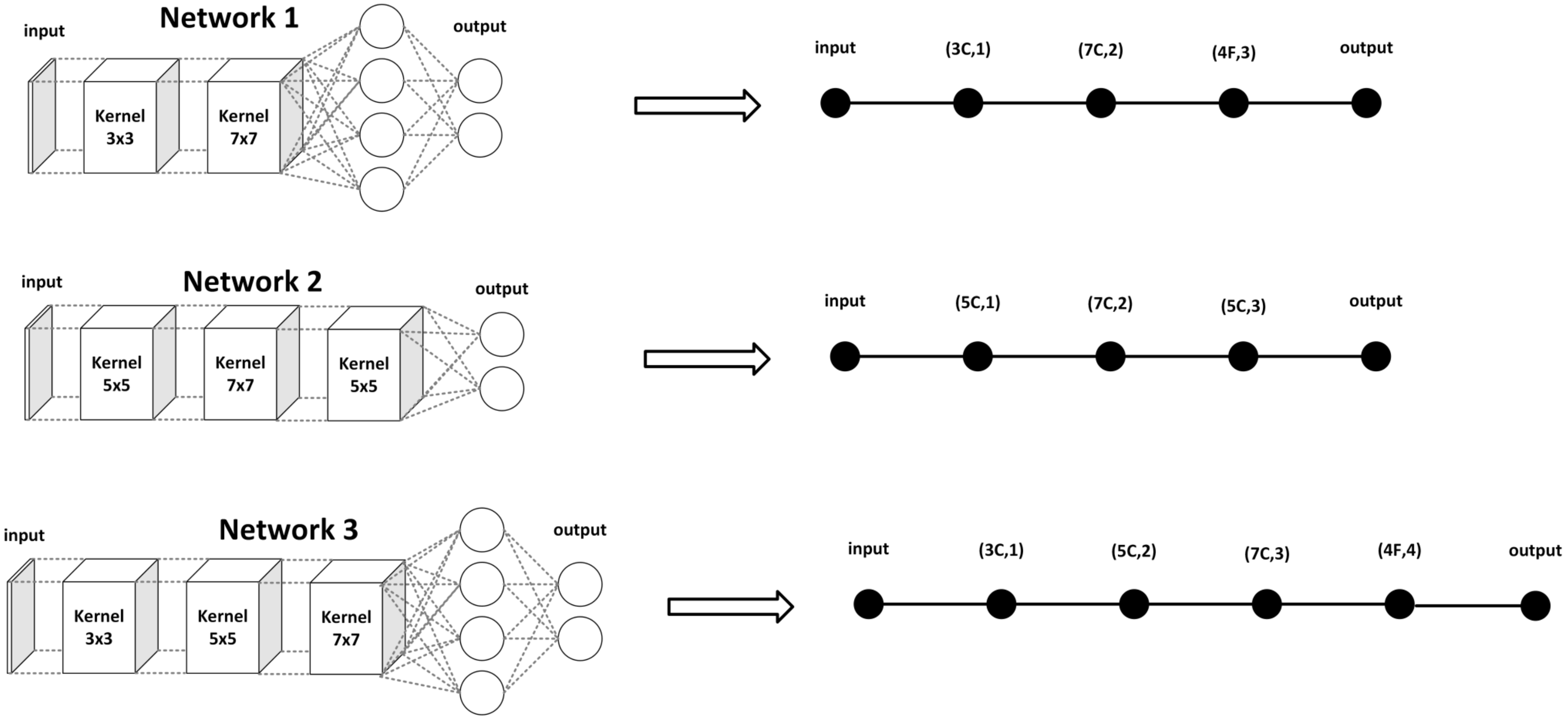
Merging the Parallel Networks

The Merging Process

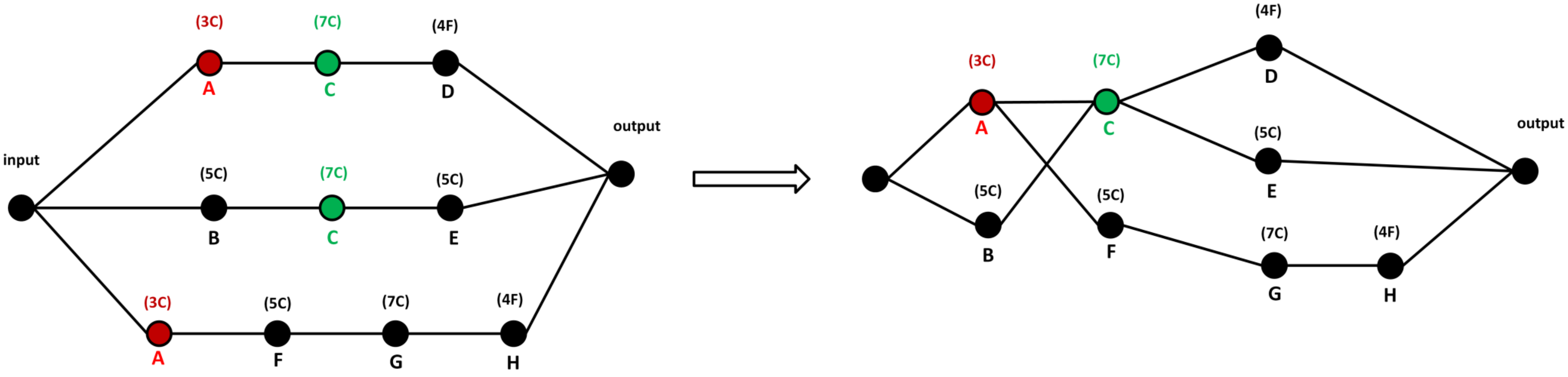
- Initially we did this intuitively, but then realized that it was based on a classic graph contraction methodology …
- We look for commonalities in the networks structures and combine these, eliminating redundancies.



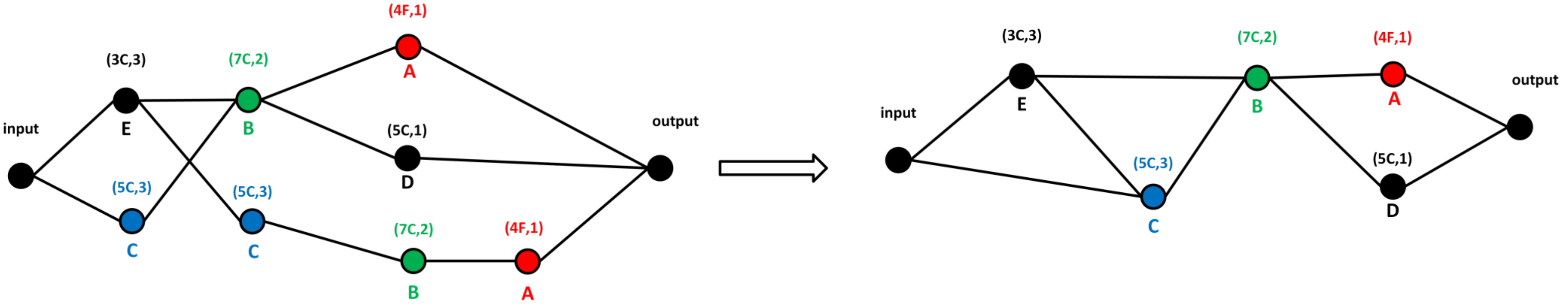
Simple Example – Networks for Iris Segmentation #2



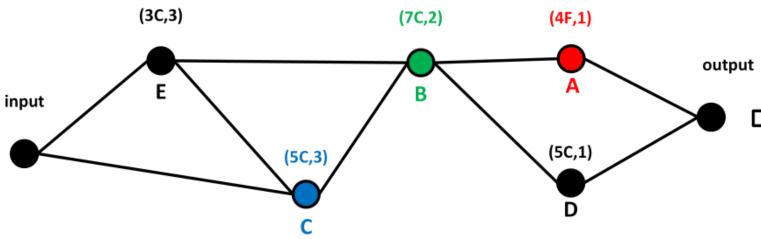
Forward Optimization, start from input node …

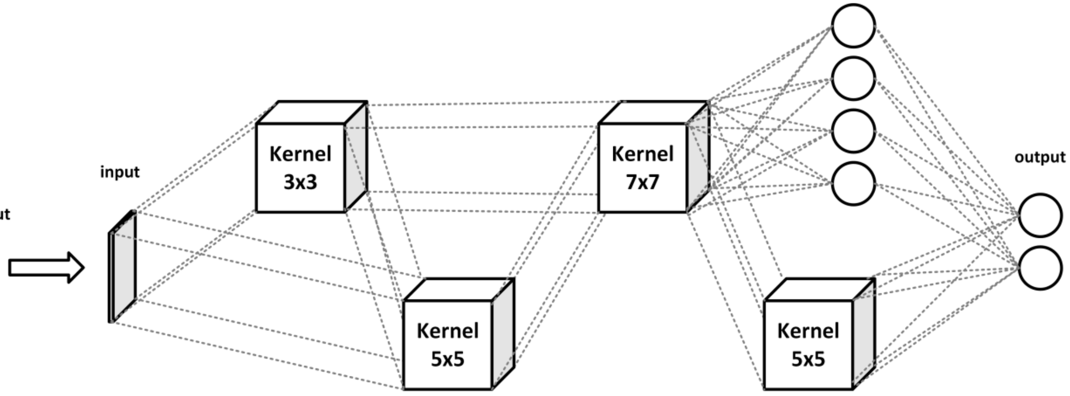


Reverse Optimization – move back from Output Node …

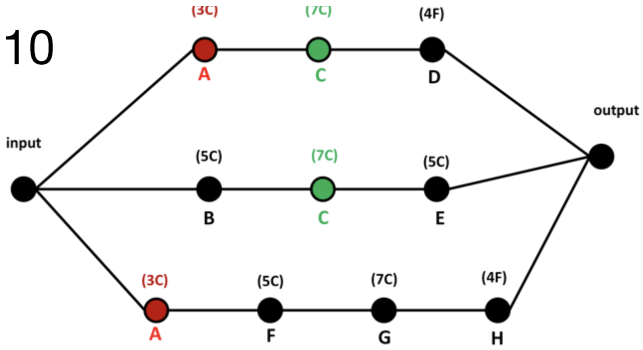


Final Network is much simplified





5 hidden layers Vs 10



Overview of The Presentation

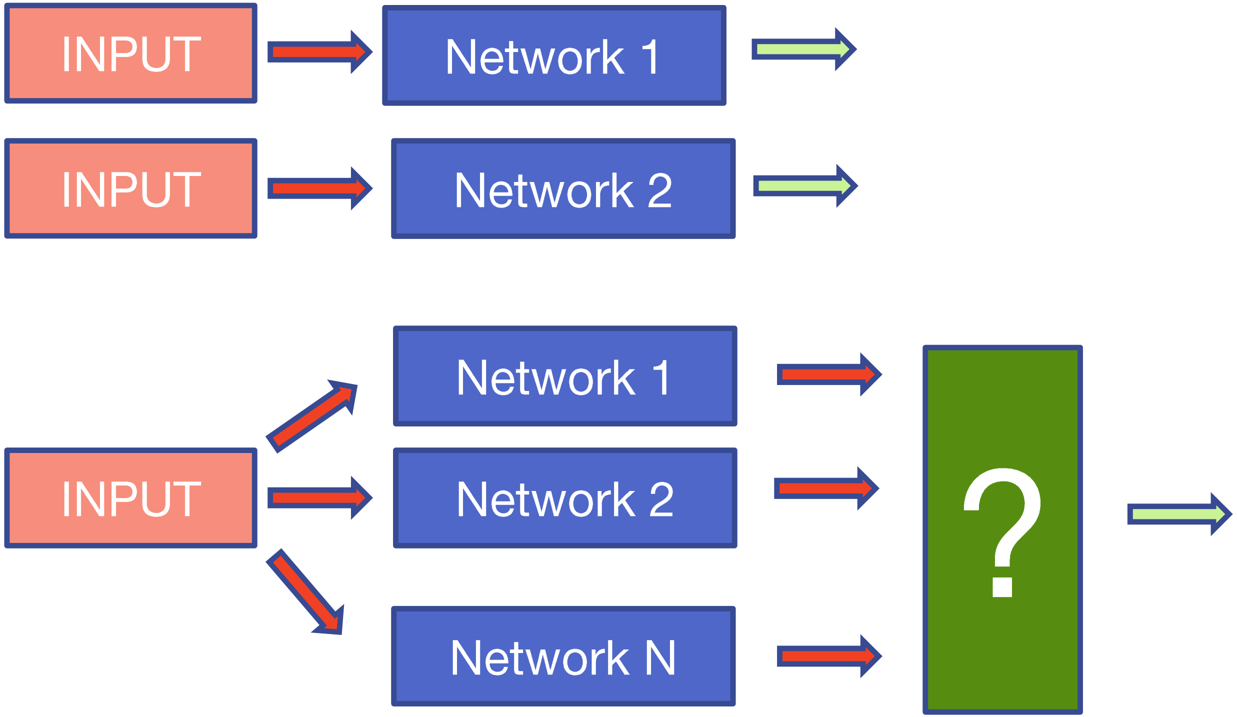
- Background & Introduction
- The Dataset Conundrum (*the back-story*)
  - Case Study – Facial Expression Datasets
- Hybrid Semi-Parallel Deep Neural Networks (SPDNN)
  - Merging the Parallel Networks
  - **Some Case Studies**
    - **Case Study #1: Iris Segmentation**
    - Case Study #2: MonoVision Depth-Map

Some Example Case Studies

First - A Comment on Metrics for combining DNNs

We can operate several individual network of size 'M' & use an elementary vote-taker or simple logic

Or we can run N networks of size M, and train a supplemental machine learning method to combine the outputs ...



A Comment on SPDNN & Metrics #2

Now if we train a SPDNN built from a set of original networks:

(i) It is always smaller in size than the total of individual networks

(ii) It does not require the additional ML block to combine the outputs from each network

(iii) The results are better than any individual network

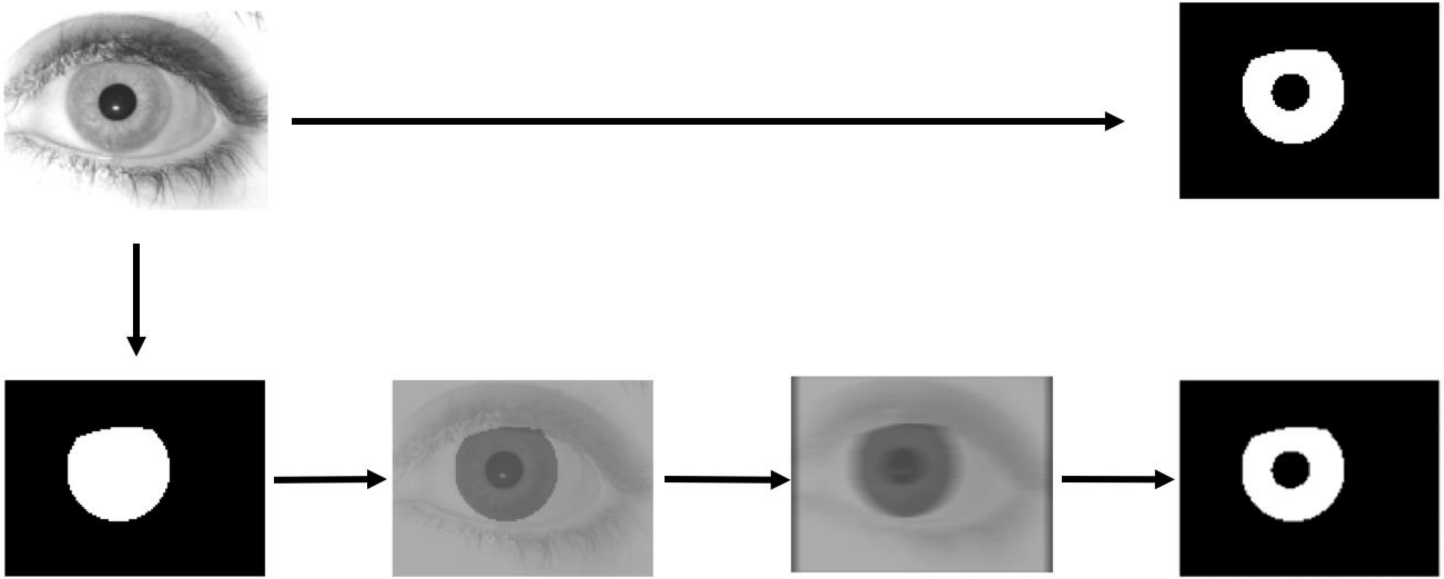
Some size metrics are given later for individual cases studies

NB: Not every SPDNN will converge; some trial & error was needed in all the cases studies given here;

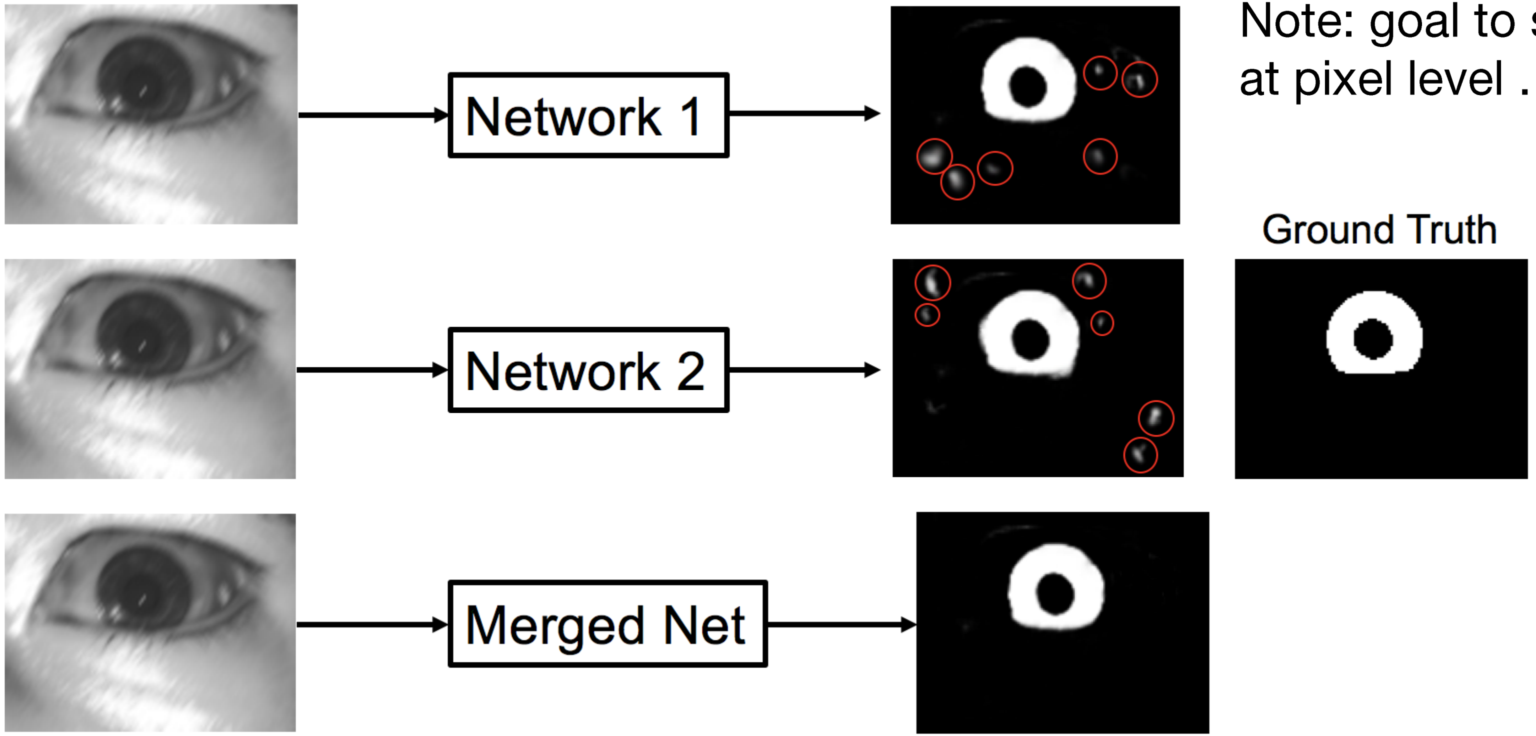
Case Study #1 – Iris Segmentation

Example – Network for Iris Segmentation

- Segmentation is a key step in iris biometrics; the bulk of errors occur due to incorrect segmentation of the iris …

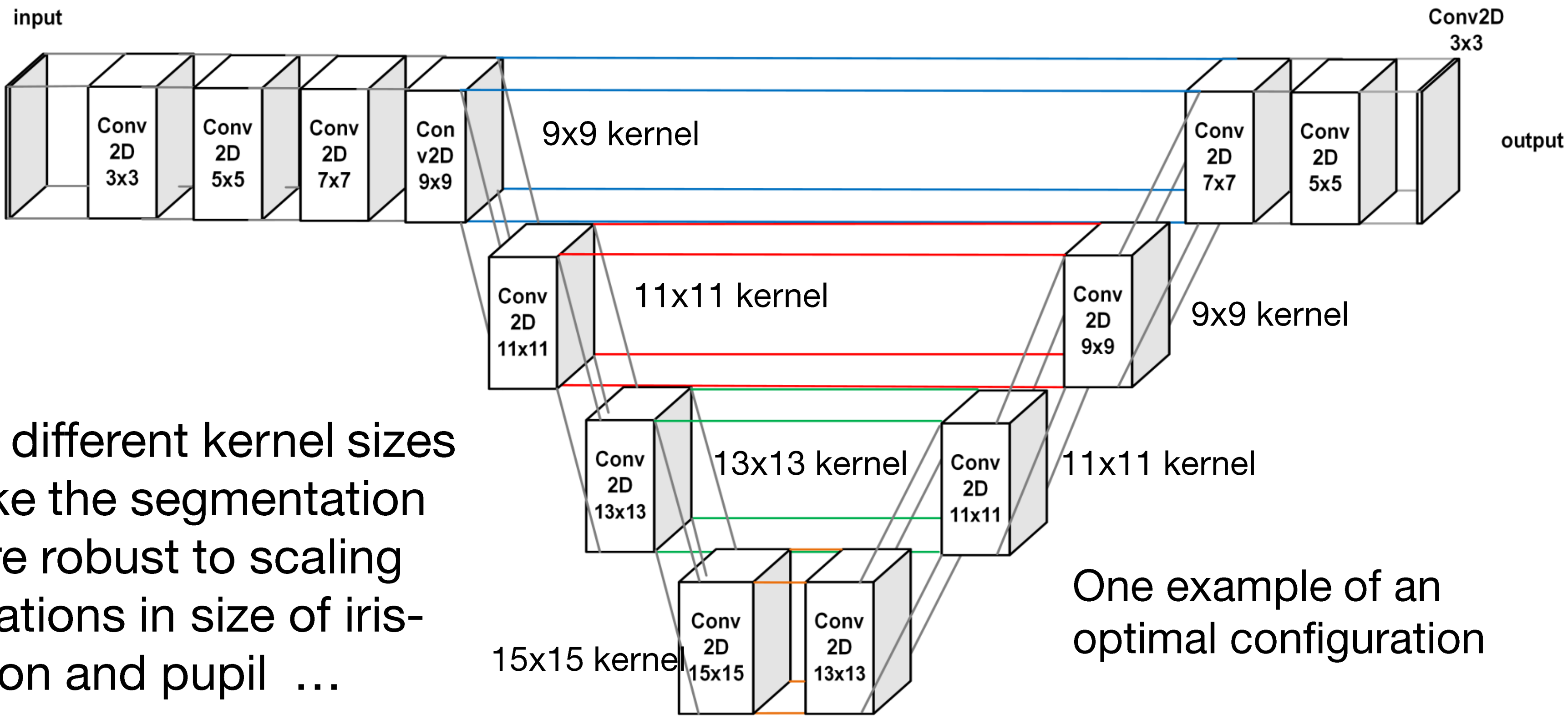


Simple DNN's have many False Positive (FP) errors



Note: goal to segment at pixel level …

Merged Hybrid SPDNN Network



- The different kernel sizes make the segmentation more robust to scaling variations in size of iris-region and pupil …

One example of an optimal configuration

Impressive Results from this Merged Network

- Table shows ***pixel-level*** segmentation accuracy
- Improvement over *state-of-art* segmentation algorithms with relatively compact network architecture:
  - Works very well on MobBio dataset - designed to be a very challenging iris segmentation dataset
  - Further optimizations – especially size reductions - have been implemented (but can't present these details) …

	UBIRIS	MobBio
IrisNet (proposed)	**99.3%**	**97.07%**
MFCN [2]	99.1%	–
HCNN [2]	98.89%	–
[11]	98.79%	–
[8]	98.69%	–
[7]	98.28%	–
[4]	98.13%	–
[6]	98.1%	–
IFFP[9]	44.38%	50.58%
GST [1]	42.59%	42.21%
WAHET [10]	27.4%	44.27%
Osiris [3]	26.46%	20.08%
CAHT [5]	18.02%	28.37%

Hyrid SPDNN for Iris Segmentation Metrics

Mono Depth	#params	Hybrid SPDNN	#params
Net1	69,480		
Net2	190,680		
Net3	451,680		
Net4	934,280		
All Nets	**1,646,120**	**Merged Network**	**1,101,580**

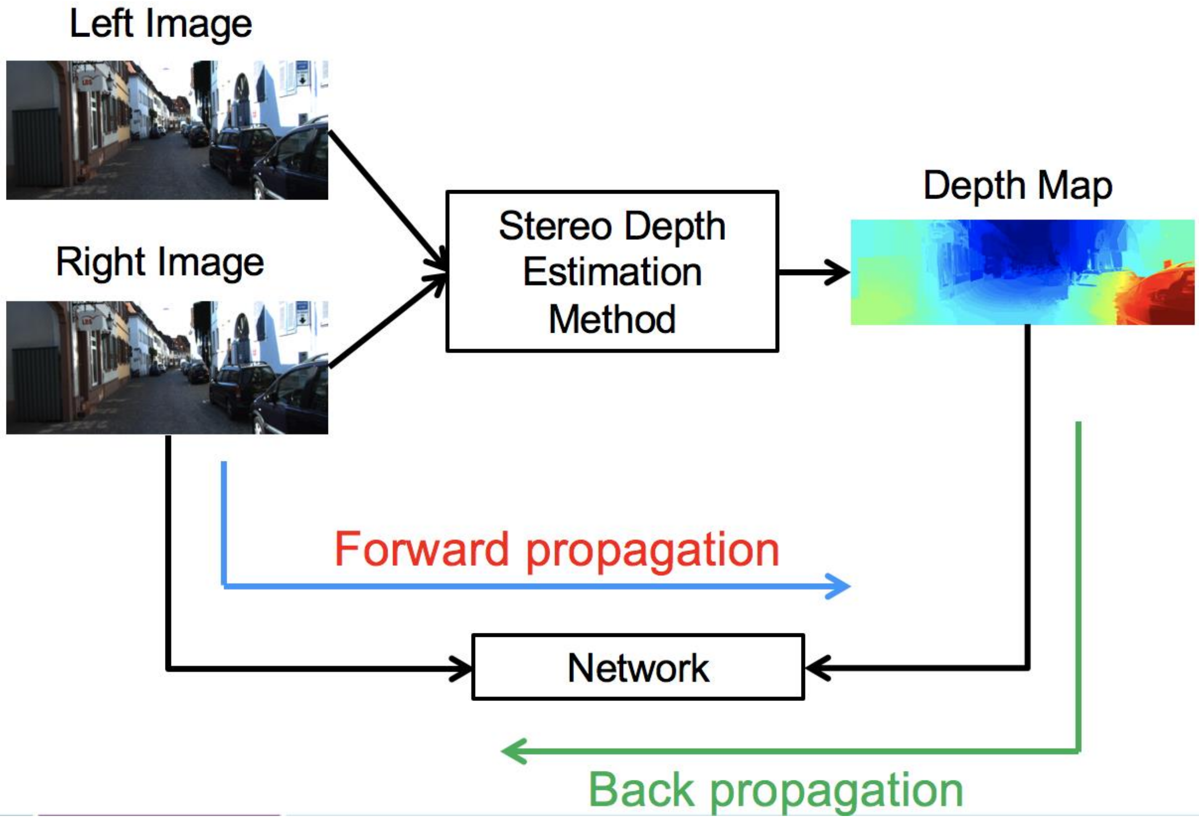
33% reduction in combined network size

Case Study #2 – MonoVision Depth-Map

Single-Camera Depth Map (for Street Scenes)

Training Methodology & Stereo Ground Truth

As an absolute ground truth is not available it is necessary to use a state-of-art stereo depth
algorithm to generate a
pseudo ground-truth



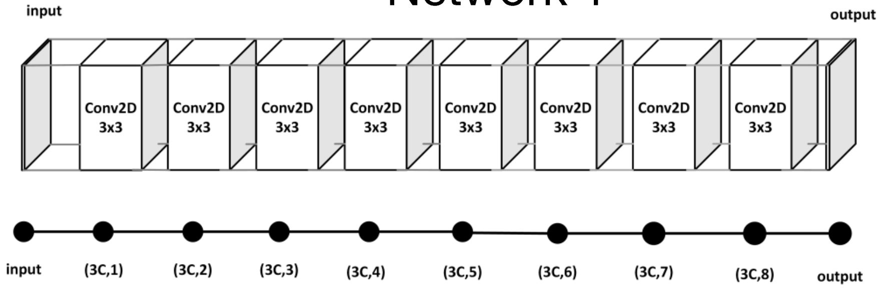
KITTI Public Dataset was used for this Research

- KITTI Stereo dataset 2012, 2015 are used for training purposes.
- The left image of the sets along with the estimated disparity map is used to train the network.
- In total 33,096 images are used in this research.
- 70% of the initial set is considered for training, 20% for validation and 10% for test purposes.

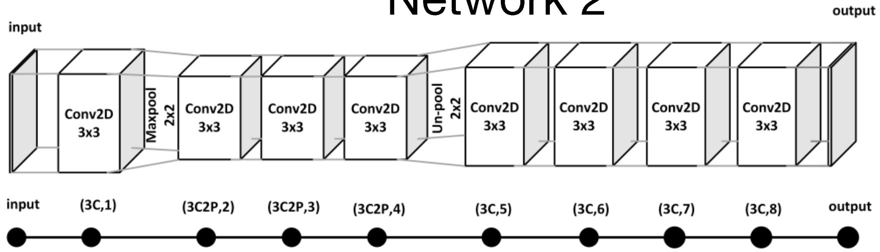
Four Fully Convolutional Networks …

- This network combines 8 distinct individual networks into a merged structure!

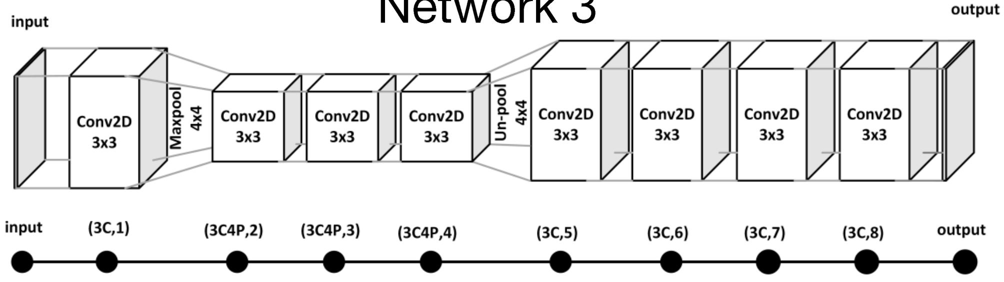
Network 1



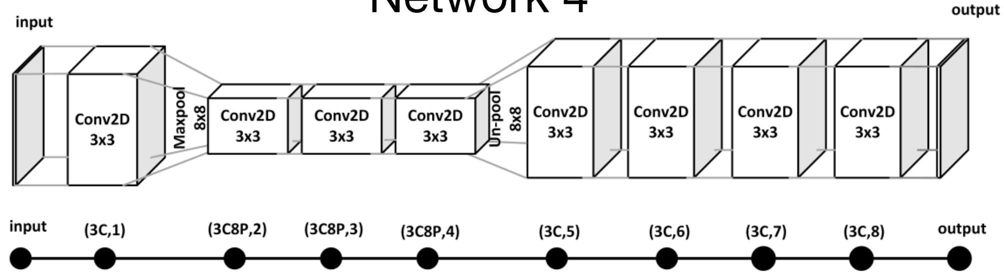
Network 2



Network 3



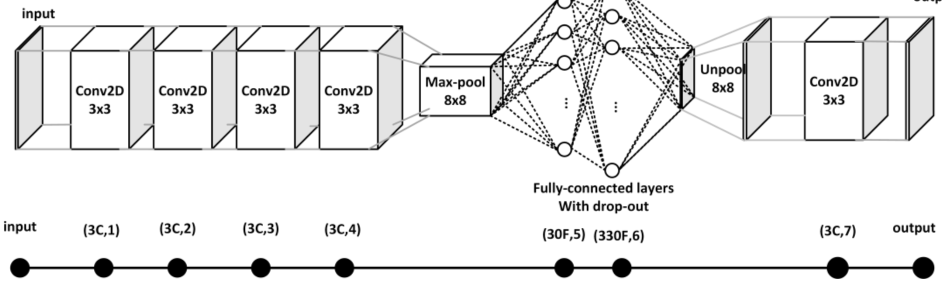
Network 4



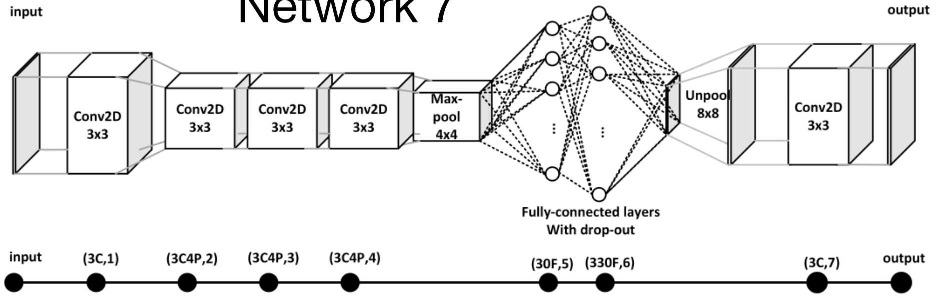
… combined with 4 Convolutional/Dense Networks

- This network combines 8 distinct individual networks into a merged structure!

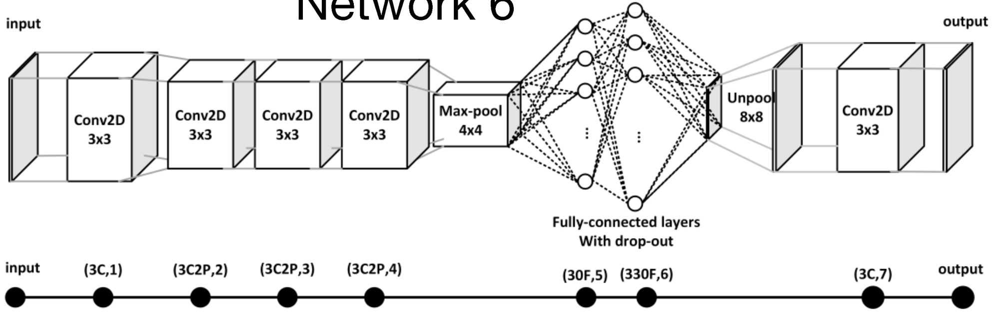
Network 5



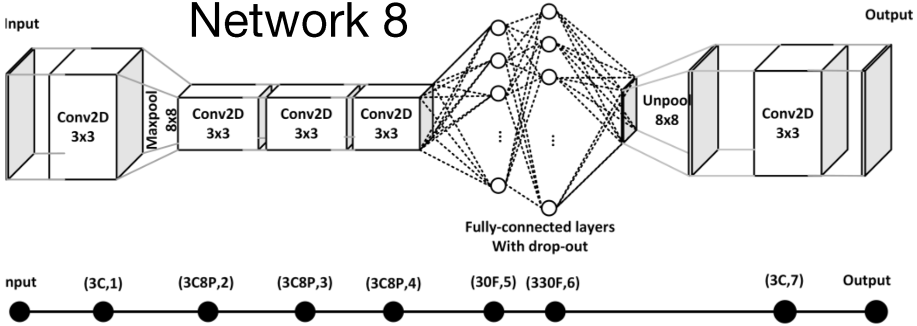
Network 7



Network 6

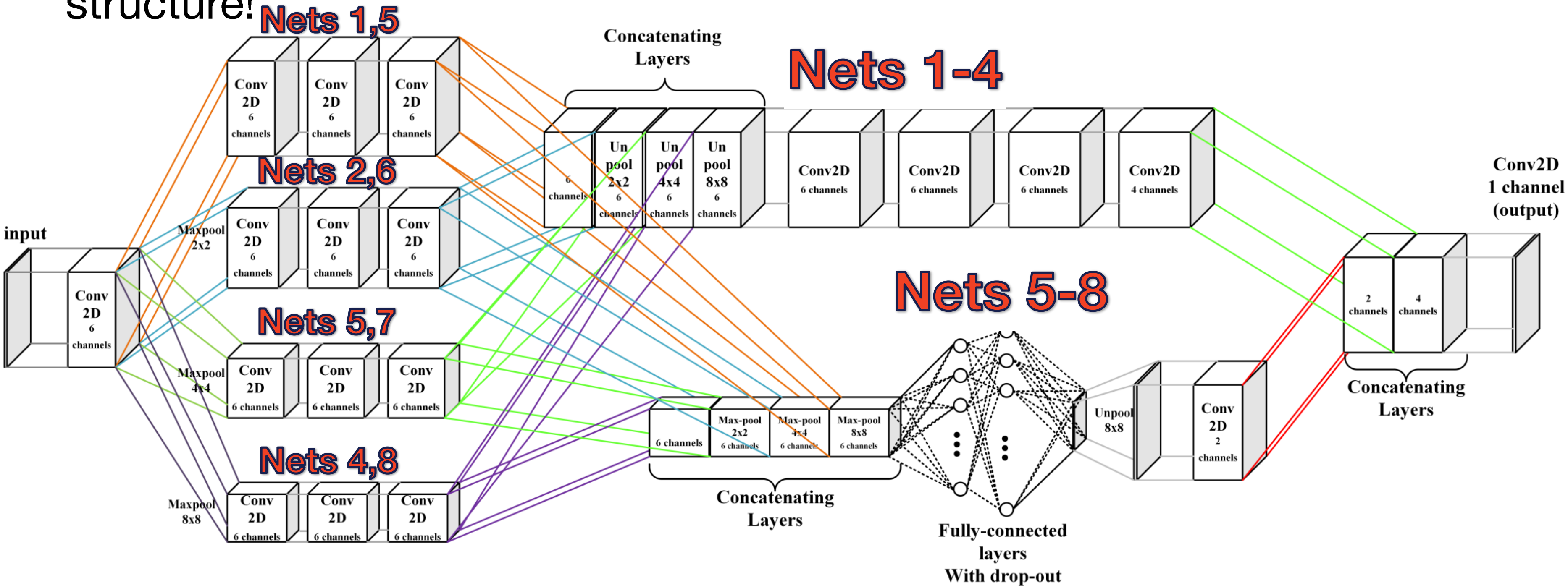


Network 8



To give … Hybrid Semi-Parallel Deep Neural Network

- This network combines 8 distinct individual networks into a merged structure!



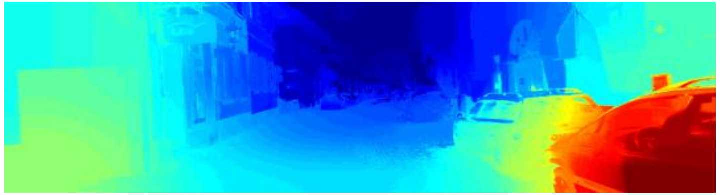
Hybrid SPDNN for Depth - Metrics

Mono Depth	#params	Hybrid SPDNN	#params
Net1	2286		
Net2	2286		
Net3	2286		
Net4	2286		
Net5	20898		
Net6	20898		
Net7	20898		
Net8	20898		
All Nets	**92736**	Merged Network	**25146**

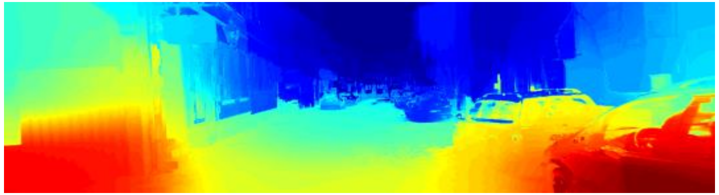
c.75% reduction in network size

Example Models (from different training subsets) #1

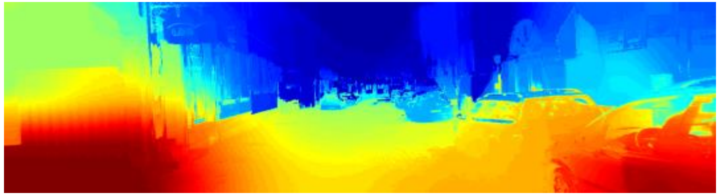
RGB Reference Frame



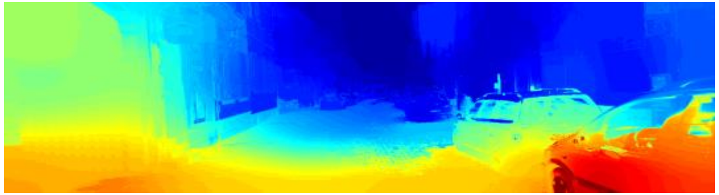
Ground Truth Computed by Stereo Matching



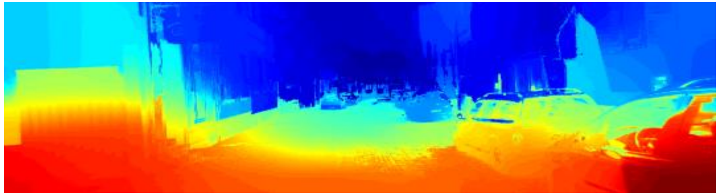
Model 1



Model 2



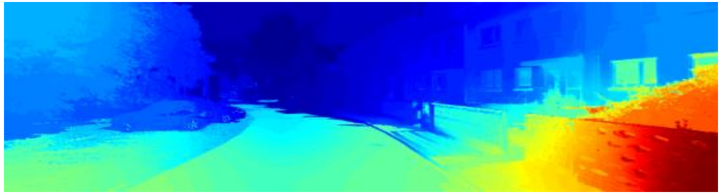
Model 3



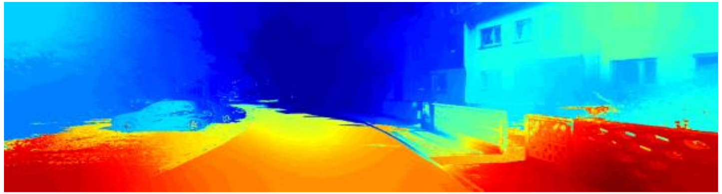
Model 4

Example Models (from different training subsets) #2

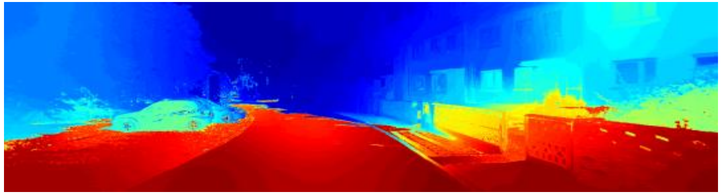
RGB Reference Frame



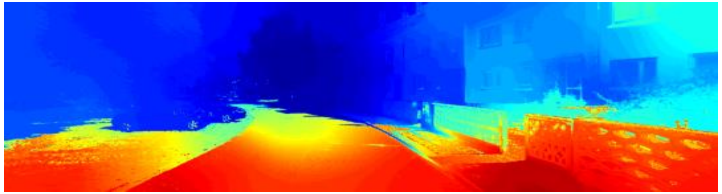
Ground Truth Computed by Stereo Matching



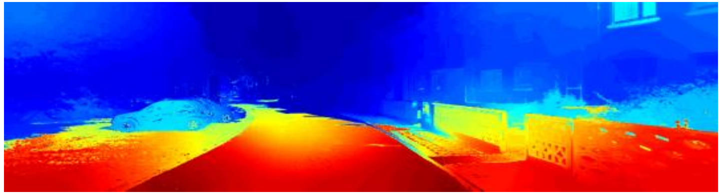
Model 1



Model 2



Model 3



Model 4

Questions?

Resource Slides

Resource Slide #1

- **Introductory Articles**
- S. Bazrafkan, T. Nedelcu, P. Filipczuk, and P. Corcoran, “Deep learning for facial expression recognition: A step closer to a smartphone that knows your moods,” in 2017 IEEE International Conference on Consumer Electronics, ICCE 2017, 2017.
- J. Lemley, S. Bazrafkan, and P. Corcoran, “Deep Learning for Consumer Devices and Services: Pushing the limits for machine learning, artificial intelligence, and computer vision.,” IEEE Consum. Electron. Mag., vol. 6, no. 2, pp. 48–56, 2017.
- S. Bazrafkan and P. M. Corcoran, “Pushing the AI Envelope: Merging Deep Networks to Accelerate Edge Artificial Intelligence in Consumer Electronics Devices and Systems,” IEEE Consum. Electron. Mag., vol. 7, no. 2, 2018.

Resource Slide #2

- **Iris Segmentation**
- S. Bazrafkan and P. Corcoran, “Enhancing Iris Authentication on Handheld Devices Using Deep Learning Derived Segmentation Techniques,” in IEEE International Conference on Consumer Electronics, (ICCE 2018), 2018.
- S. Bazrafkan, S. Thavalengal, and P. Corcoran, “An End to End Deep Neural Network for Iris Segmentation in Unconstraint Scenarios,” arXiv Prepr. arXiv1712.02877, 2017.

- **MonoVision Depth**
- Bazrafkan S, Javidnia H, Lemley J, Corcoran P. Depth from Monocular Images using a Semi-Parallel Deep Neural Network (SPDNN) Hybrid Architecture. arXiv preprint arXiv:1703.03867. 2017 Mar 10.

Resource Slide #3 – Datasets

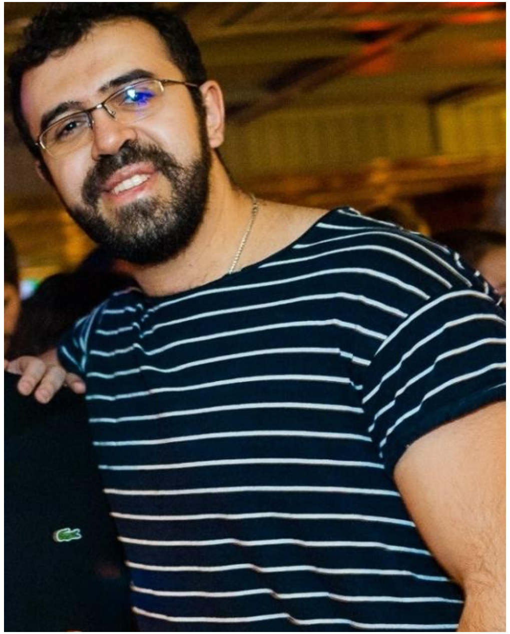
- JAFFE Dataset: http://www.kasrl.org/jaffe.html
- The Radboud Faces Database (RaFD)
  - http://www.socsci.ru.nl:8180/RaFD2/RaFD?p=main
- Cohn-Kanade (CK and CK+) database
  - http://www.consortium.ri.cmu.edu/ckagree/
- KITTI Vision Benchmark Suite: http://www.cvlibs.net/datasets/kitti/
- MobBio - https://paginas.fe.up.pt/~mobbio2013/
- UBIRIS - http://iris.di.ubi.pt/

Back-Up Slides

The originator of Hybrid Semi-Parallel Networks

- Shabab Bazrafkan

Shabab received his B.Sc degree from Urmia University, Urmia, Iran in electrical engineering in 2011 and M.Sc degree from Shiraz University of Technology (SuTECH) in telecommunication engineering, Image processing branch in 2013. Currently he is a PhD student with Cognitive, Connected & Computational Imaging Research group at the National University of Ireland, Galway (NUIG); his research work is funded via an Academic/Industry partnership jointly sponsored by Science Foundation Ireland (SFI) and industry partner, FotoNation Ltd. His main field of research is Deep Neural Networks and Neural Network design applied to a range of problems in computer vision and *Next Generation Smartphone Imaging*.



Other Works in Progress

De-Mosaicing of Noisy Images

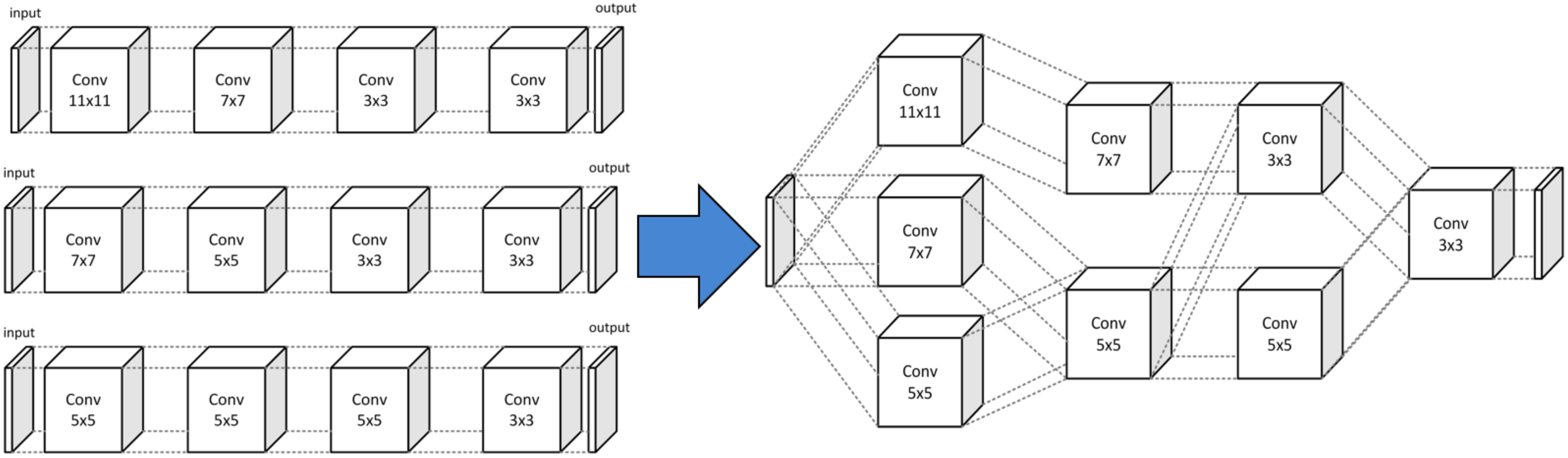
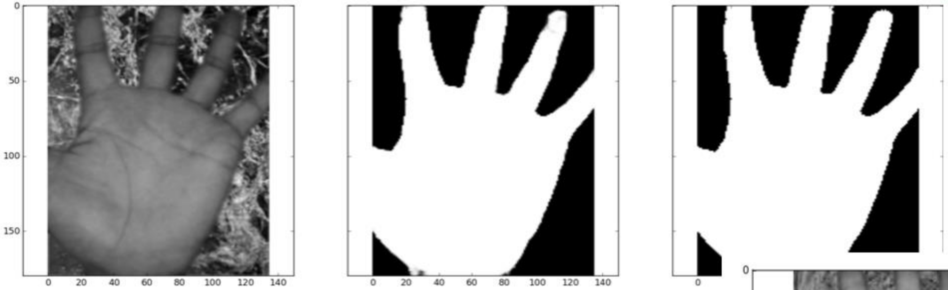


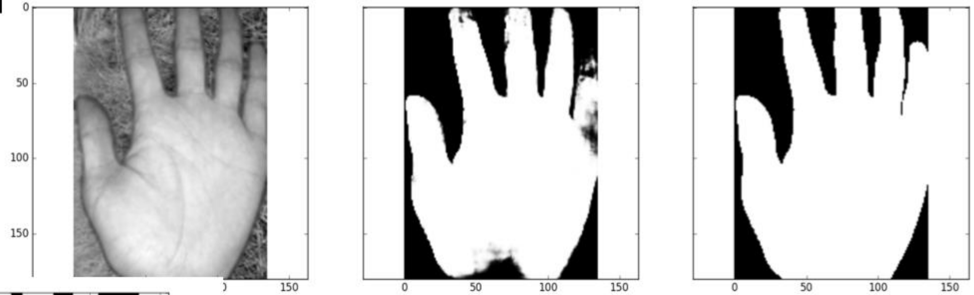
Image De-Mosaicing – some Initial Results

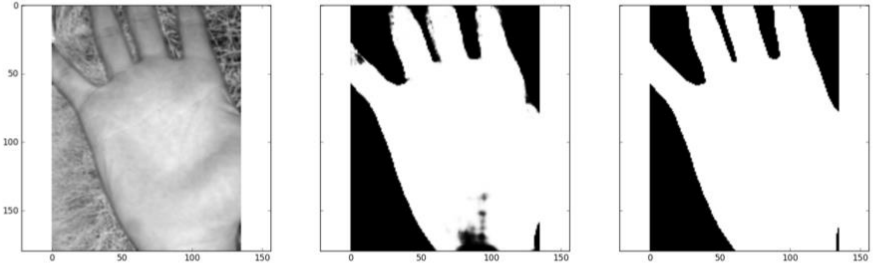
Data generated with different noise levels (б x 10^{-4})	Model trained using data with different noise levels (б x 10^{-4})						
	б = 0	б = 25	б = 50	б = 75	б = 100	б = 125	б = 150
б = 0	50.191	49.454	48.354	46.927	46.739	44.782	45.502
б = 25	46.703	48.530	48.123	46.854	46.673	44.818	45.497
б = 50	43.554	46.355	47.338	46.488	46.463	44.832	45.455
б = 75	41.437	44.230	45.955	45.726	46.086	44.750	45.322
б = 100	39.861	42.434	44.228	44.637	45.521	44.545	45.064
б = 125	38.602	40.907	42.440	43.361	44.773	44.181	44.694
б = 150	37.544	39.577	40.763	42.023	43.853	43.616	44.228

	Model trained using data with different noise levels (б x 10^{-4})						
	б = 0	б = 25	б = 50	б = 75	б = 100	б = 125	б = 150
PSNR score	53.7767	47.9876	43.7176	40.8200	38.6615	36.9561	35.5571

Hand Segmentation for Palmprint Recognition







		SPDNN9	SegNet basic
Accuracy	mean	**99.7 %**	99.51 %
	var	0.6 %	**0.57 %**
Sensitivity	mean	**99.72 %**	99.55 %
	var	0.54 %	**0.45 %**
Specificity	mean	**99.65 %**	99.42 %
	var	**0.88 %**	0.92 %
Precision	mean	**99.72 %**	99.54 %
	var	**0.66 %**	0.69 %
NPV	mean	**99.66 %**	99.43 %
	var	0.7 %	**0.63 %**
F1Score	mean	**99.72 %**	99.55 %
	var	0.54 %	**0.51 %**
MCC	mean	**99.38 %**	98.98 %
	var	1.19 %	**1.12 %**
Informedness	mean	**0.9937**	0.9898
	var	0.0123	**0.0116**

		SPDNN9	SegNet basic
FPR	mean	**0.34 %**	0.57 %
	var	**0.88 %**	0.92 %
FNR	mean	**0.27 %**	0.44 %
	var	0.54 %	**0.45 %**
FDR	mean	**0.27 %**	0.45 %
	var	**0.66 %**	0.69 %