embedded VISION SUMMIT 2018

Implementing and Improving Traditional Computer Vision Algorithms using DNN Techniques

Imagination

May 23, 2018

Overview

- Introduction
- Mapping traditional vision to deep learning primitives
- Traditional vision to learning-based vision

Imagination: A global technology leader

A technology powerhouse for multimedia and communications IP

Developing innovative IP

- PowerVR GPU: leader in graphics & compute
- PowerVR Vision & AI: dedicated IP for vision & AI
- Ensigma: connectivity and broadcast communications IP

Delivering exceptional service

- Enabling very fast time to market
- Enabling customers to leverage IP to maximise differentiation

Driving major markets

- Helping our partners to create successful solutions
- Influencing new and emerging opportunities
- Showcasing and proving our technology with real products

Products

- More than 10bn units shipped
- Over 3m per day
- Around 1.2bn in past year

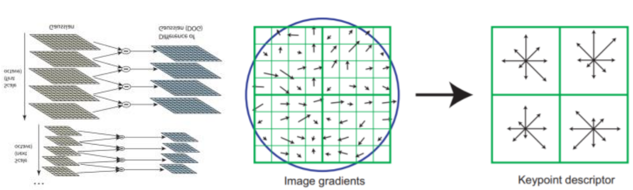
People

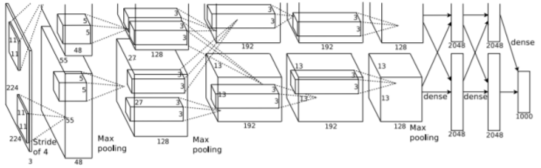
- HQ in UK
- >800 people world-wide
- >80% of staff are engineers

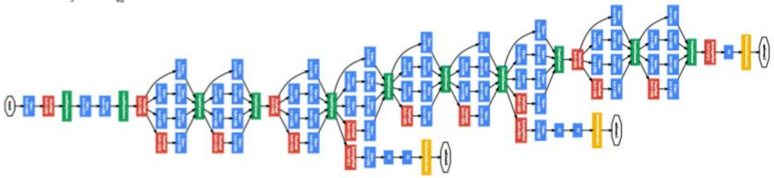
Introduction

- Traditional vision -> deep learning
- Significant investment in traditional vision
  – Expect to move over to deep learning for some or all algorithms
- Choices today when designing or selecting silicon
  – Optimise for deep learning
  – Optimise for traditional vision
  – Increased cost to provide dedicated support for both
- Programmable solutions available do not ‘solve’ the problem
  – Still need to make a choice about priorities

Traditional Vision v Deep Learning



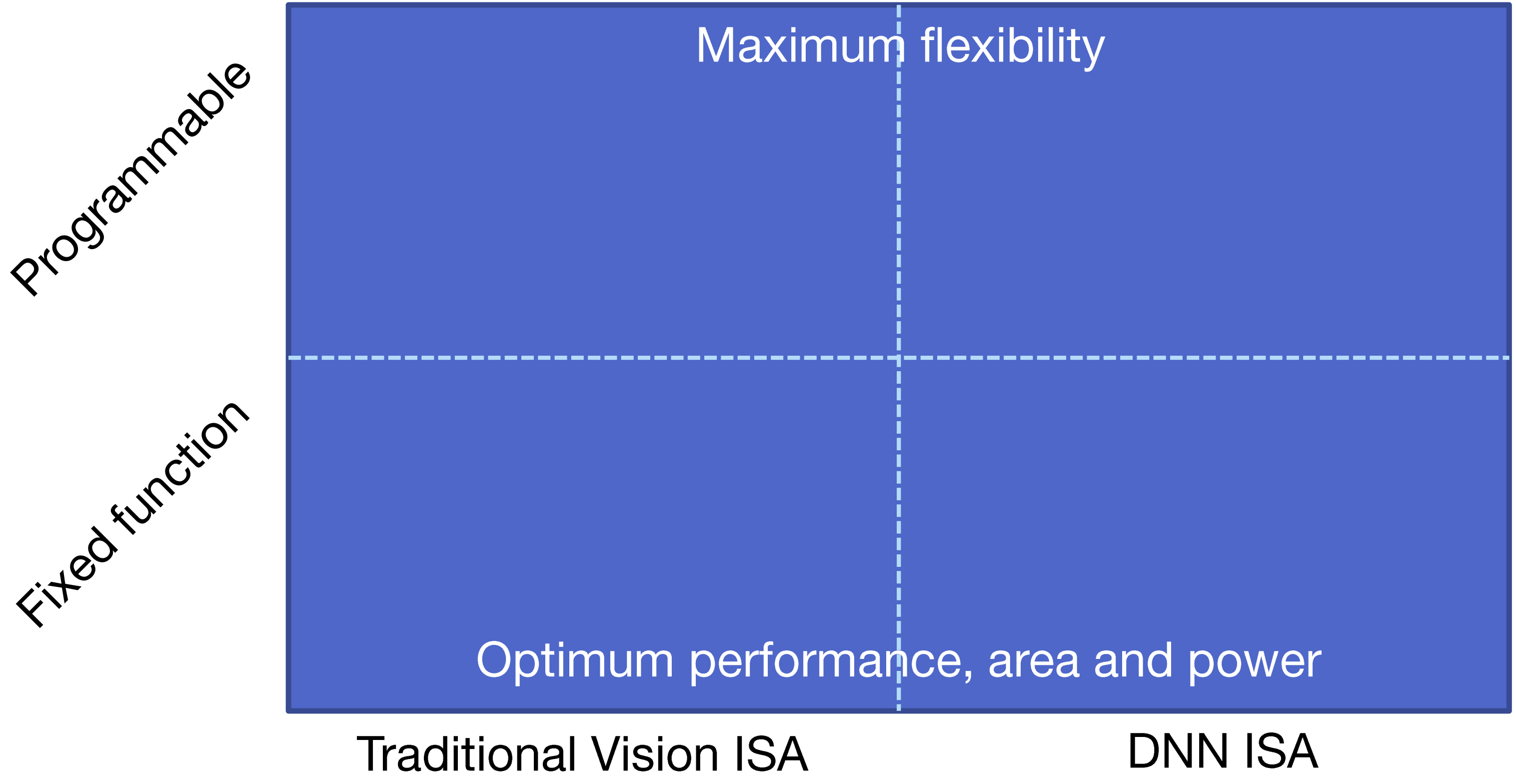




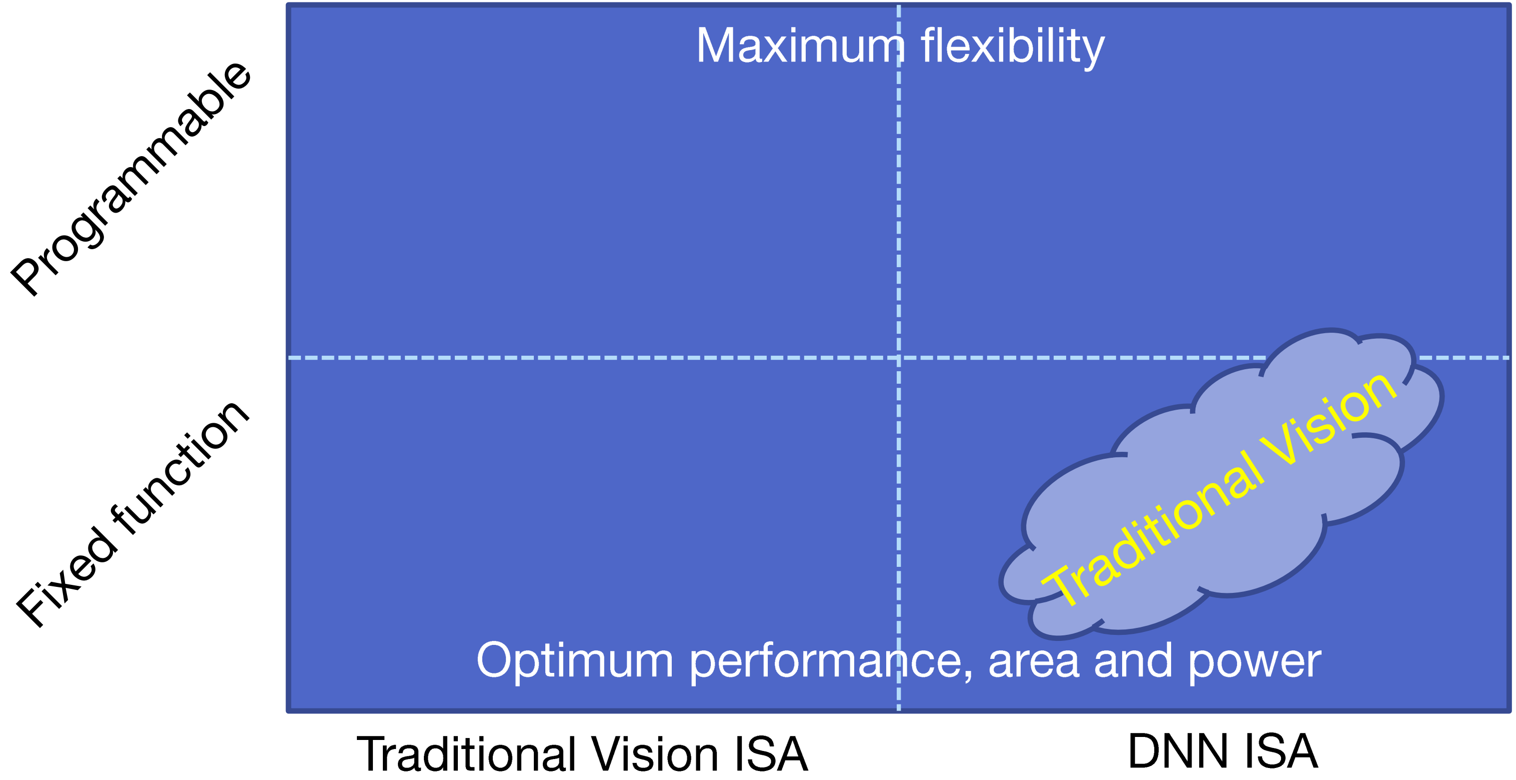
- Traditional Vision
  - > 30 years investment by industry
  - Often good enough for specific tasks
  - Proven, mature and understood
  - Diverse set of primitive operations
    - Lots of algorithmic secret-sauce

- Deep Learning
  - Huge investment over past 3-4 years
  - State-of-the-art for many tasks
  - New model development takes investment
  - Relatively small set of basic primitives
    - Although increasing

Processors for Embedded Vision



Processors for Embedded Vision



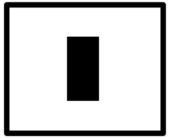
Deep Learning Primitives

- DNN primitives
  - Convolution
  - Deconvolution
  - Fully Connected
  - Activation
  - Pooling
  - Tensor Operations
  - Normalisation

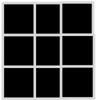
Supported by DNN processors

- Expressing traditional vision using DL primitives
  - Run on DNN processor

Binary Morphological Operator - Dilate



Input Image

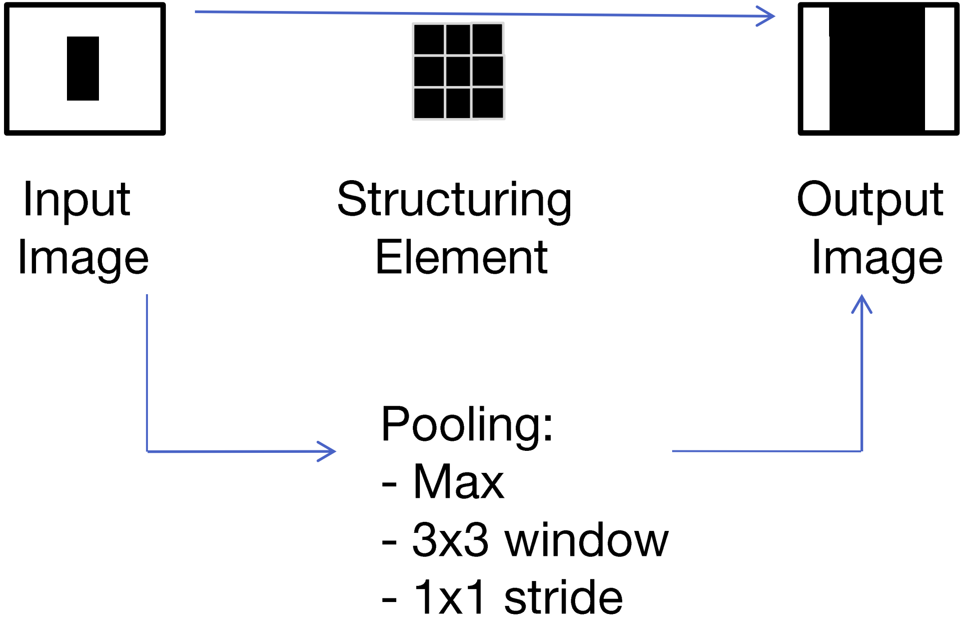


Structuring Element

Output Image

Mapping Traditional Vision to DNN Primitives

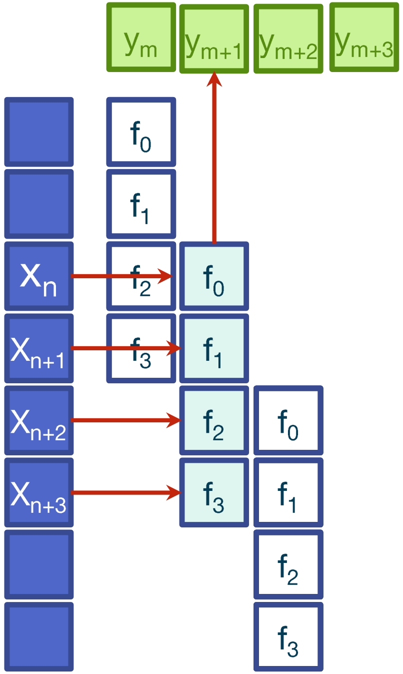
Binary Morphological Operator - Dilate



Mapping Traditional Vision to DNN Primitives

2x Bilinear Up-Scaling

- Make use of deconvolution primitives

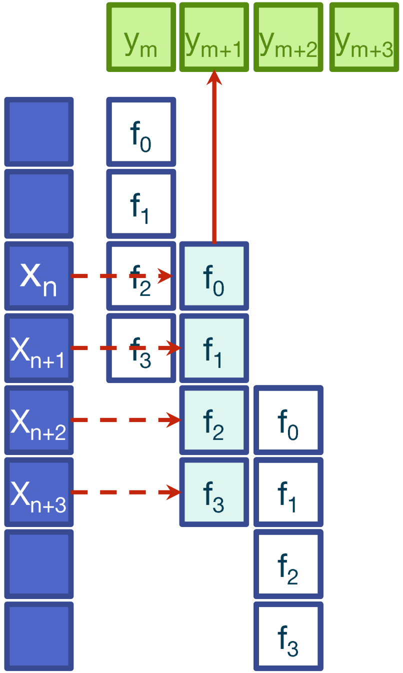


Convolution with stride = 2

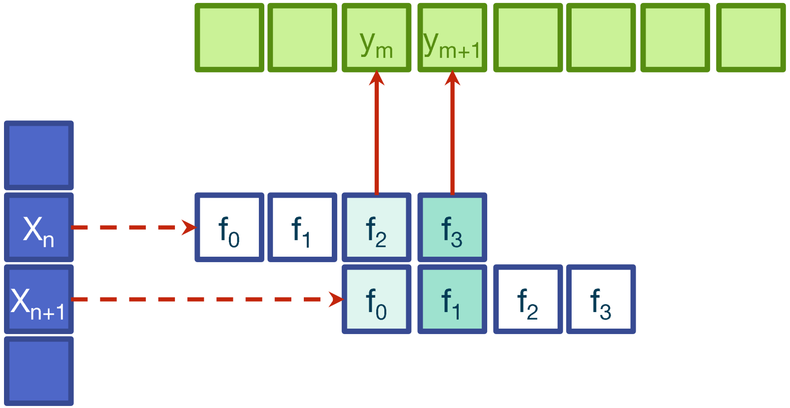
Mapping Traditional Vision to DNN Primitives

2x Bilinear Up-Scaling

- Make use of deconvolution primitive



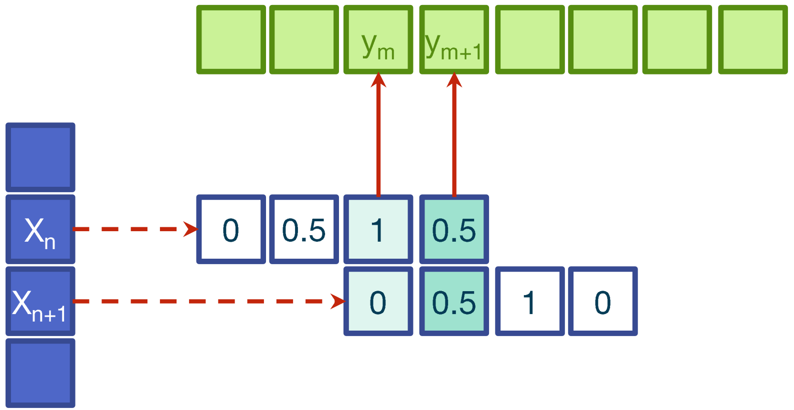
Deconvolution with scaling factor = 2



Mapping Traditional Vision to DNN Primitives

2x Bilinear Up-Scaling

Linear upscale with scaling factor = 2



Mapping Traditional Vision to DNN Primitives

2x Bilinear Up-Scaling

2D linear upsample filter with scaling factor = 2

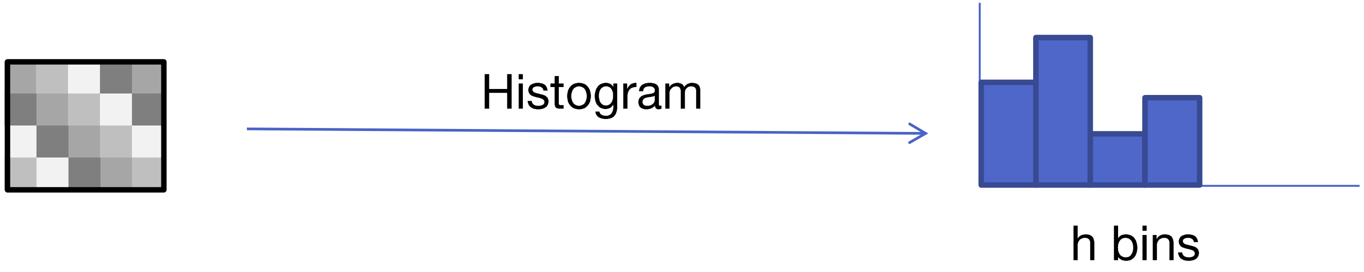
$$\begin{bmatrix} 0.25 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 0.25 \end{bmatrix}$$

$$\begin{bmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{bmatrix} \quad \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 0.5 & 0.5 \\ 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

Mapping Traditional Vision to DNN Primitives

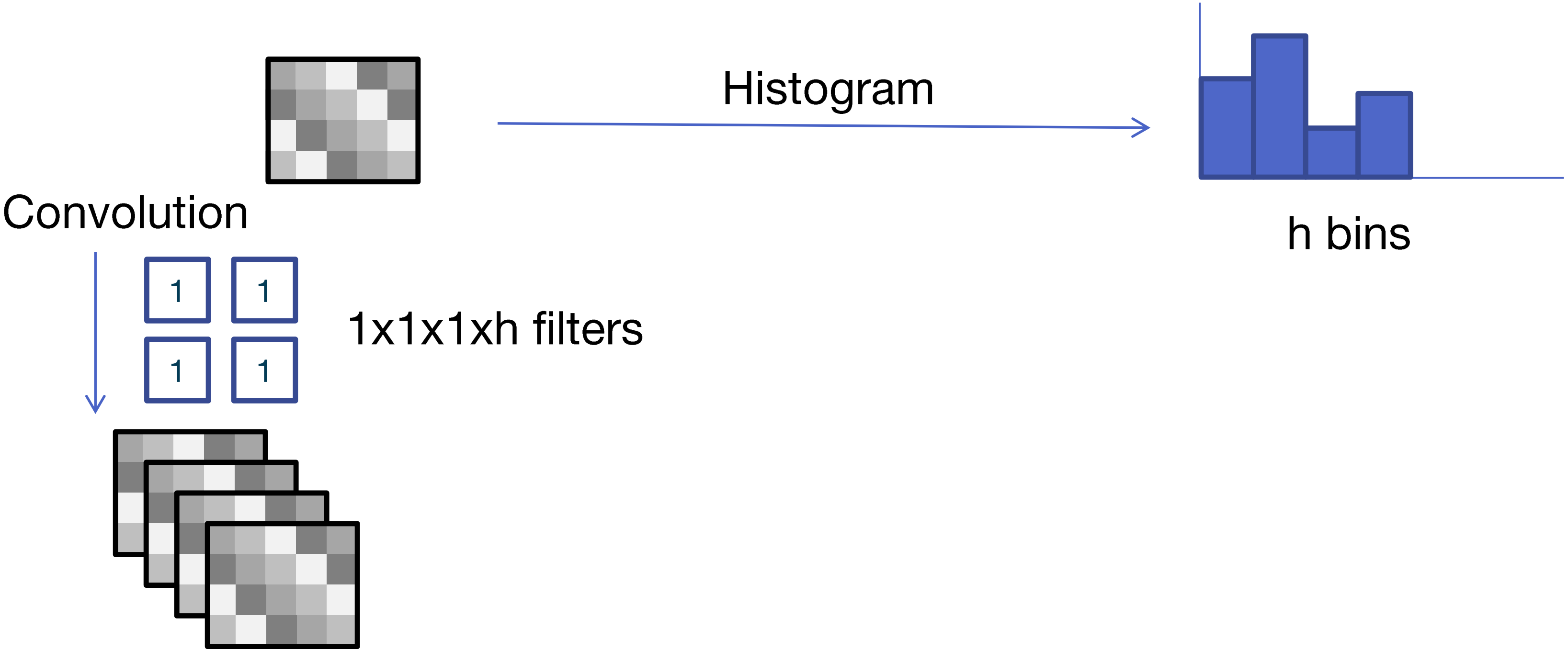
Histogram



Mapping Traditional Vision to DNN Primitives

Histogram

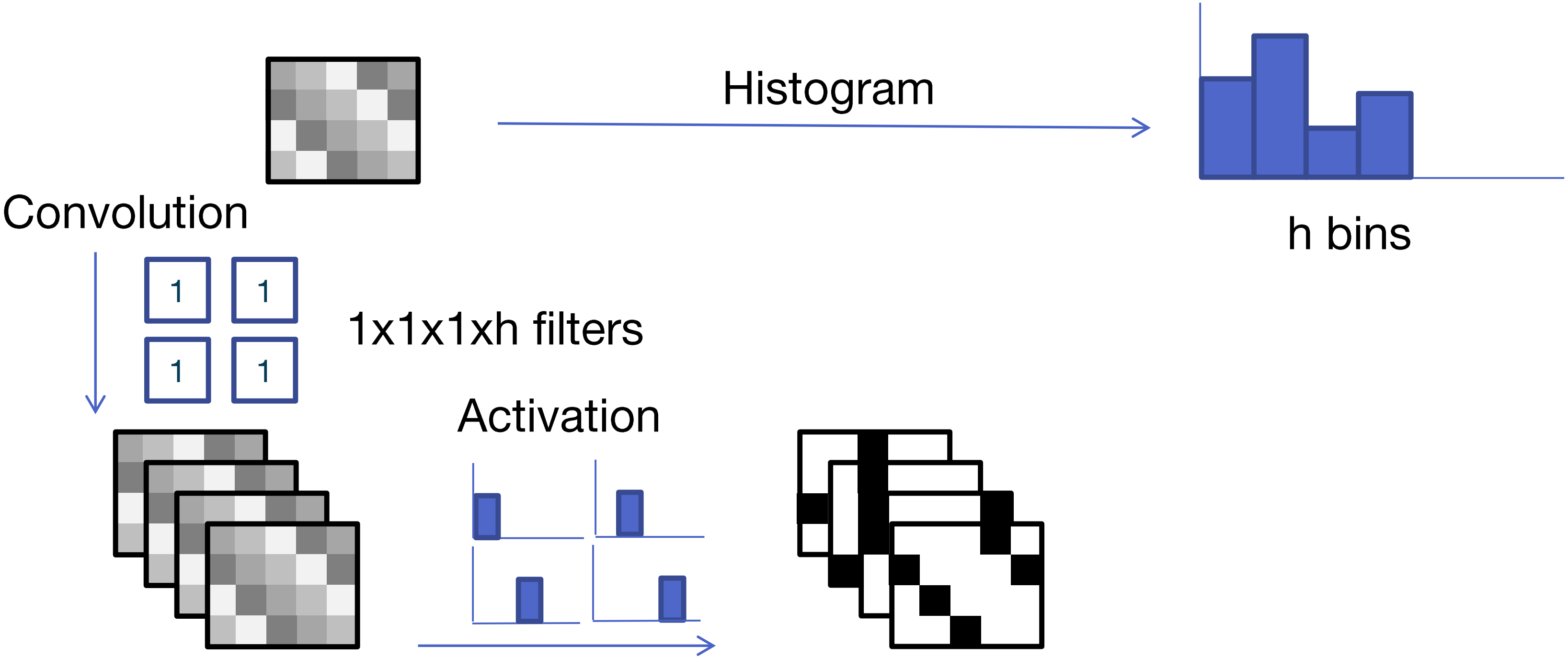
- Convolution, activation and pooling



Mapping Traditional Vision to DNN Primitives

Histogram

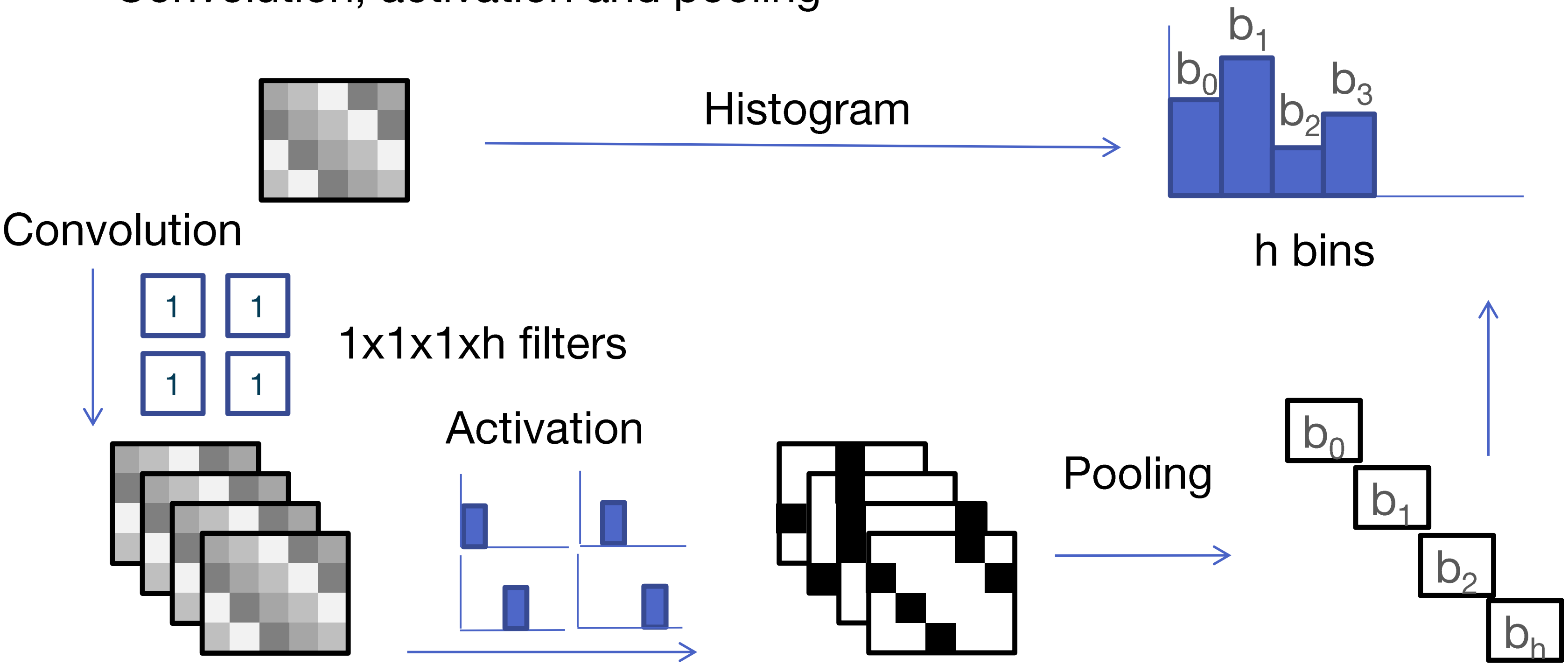
- Convolution, activation and pooling



Mapping Traditional Vision to DNN Primitives

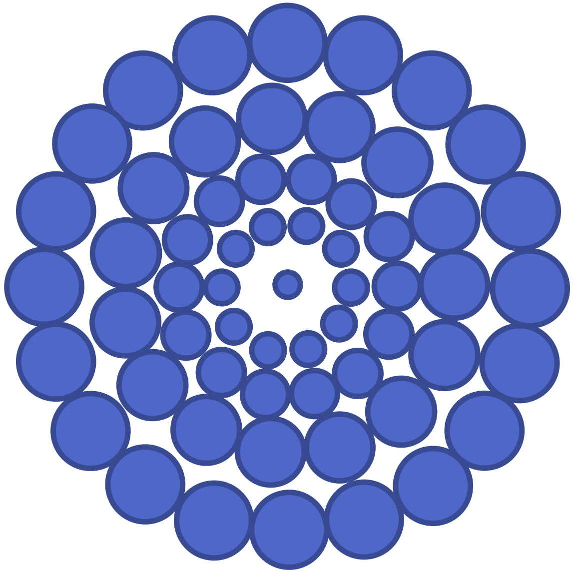
Histogram

- Convolution, activation and pooling



Mapping Traditional Vision to DNN Primitives

Feature Descriptor: Brisk



BRISK sampling positions

- Well-known & widely used binary image feature descriptor
- Good example of traditional vision
  - Hand designed, based on insights and intuition
- Defined 60 samples around feature point
- Set of 512 ‘short-distance’ pairs defined S

$$B = \{b_{ij}\}, b_{ij} = \begin{cases} 1, I(p_i) > I(p_j) \\ 0, \text{otherwise} \end{cases}$$

- Generates a 512D patch descriptor

Mapping Traditional Vision to DNN Primitives

Feature Descriptor: Brisk

- Express Brisk feature pairs as (sparse) matrix

$$\boldsymbol{D} = \text{sigmoid}\left(\begin{bmatrix} 1 & 0 & \cdots & -1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & -1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & -1 \end{bmatrix} \begin{bmatrix} \boldsymbol{I}(p_{\mathbf{1}}) \\ \boldsymbol{I}(p_{\mathbf{2}}) \\ \vdots \\ \boldsymbol{I}(p_{\mathbf{60}}) \end{bmatrix}\right)$$

- Map to DNN primitives
  - Fully connected (matrix-vector multiply)
  - Activation

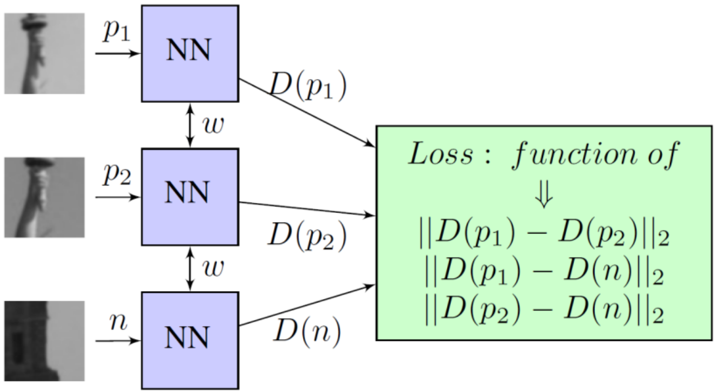
Learning a Feature Descriptor: Brisk

Learning a better descriptor

- Set up a trainable weight matrix

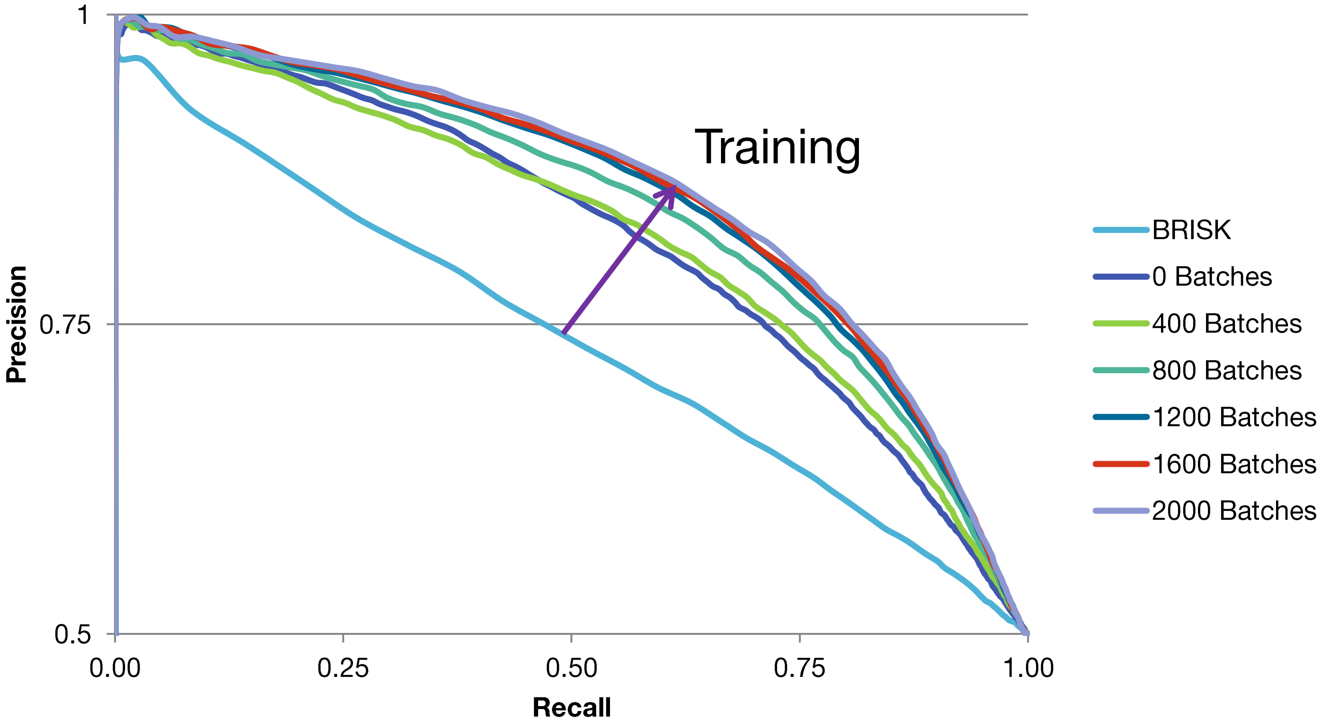
$$\boldsymbol{D} = \text{sigmoid}\left(\begin{bmatrix} w_{1,1} & \cdots & w_{1,60} \\ \vdots & \ddots & \vdots \\ w_{512,1} & \cdots & w_{512,60} \end{bmatrix} \begin{bmatrix} I(p_1) \\ I(p_2) \\ \vdots \\ I(p_{60}) \end{bmatrix} \right),$$

- Use triplet-loss to simultaneously
  - Maximise distance between different patches
  - Minimise distance between same patches
- Use HPatches dataset

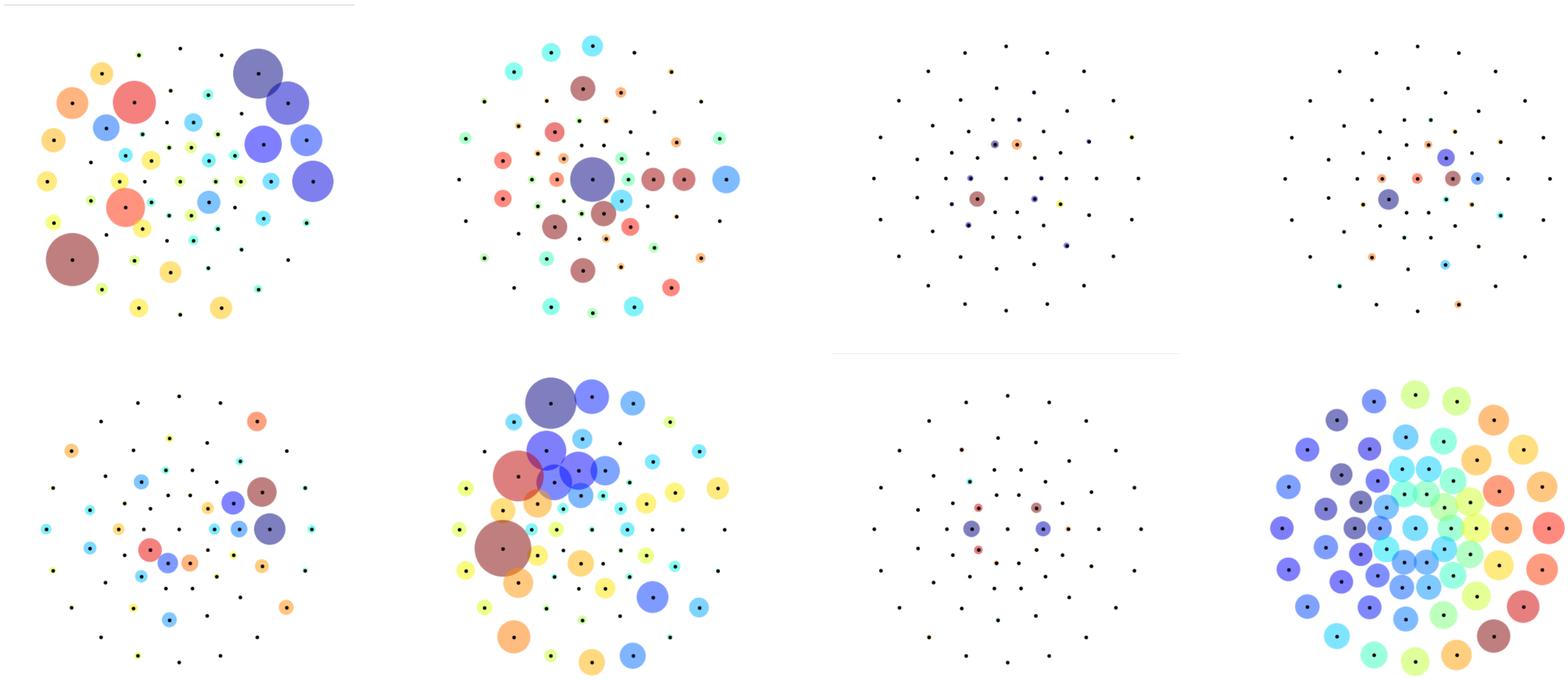


Feature Descriptor: Brisk

Results: HPatches (Easy)



Example learnt 'features'



Summary

- Deep learning defines a set of basic primitives
- Traditional computer vision can be enabled in a deep learning world by exploiting these primitives
- Traditional vision algorithms can then be used as a starting point for learning
  - Leverage existing investment
  - Provides a path to take incremental steps into learning

Resources

- PowerVR Blog
  - https://www.imgtec.com/blog/
- Traditional Vision
  - Computer Vision: Algorithms and Applications (Richard Szeliski)
  - http://szeliski.org/Book/
- Other Reading
  - http://zbigatron.com/has-deep-learning-superseded-traditional-computer-vision-techniques/