Sketch-a-Net that Beats Humans

Published in: British Machine Vision Conference 2015

Authors: Qian Yu, Yongxin Yang
Yi-Zhe Song, Tao Xiang, Timothy Hospedales

Centre for Intelligent Sensing
Queen Mary University of London

CIS centre for intelligent sensing

Queen Mary University of London

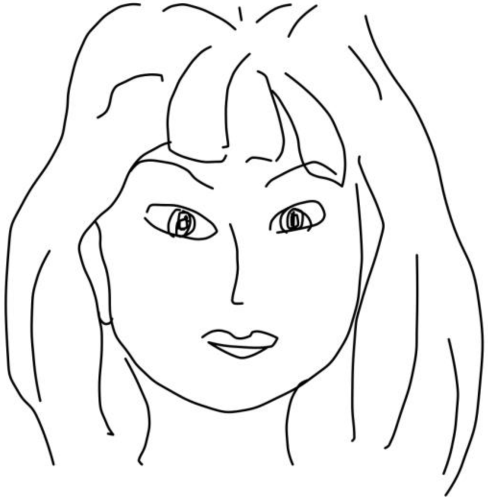
Overview

- Background
- Methodology
- Results

Free-hand sketch

- Definition: drawn by **non-artists** on touchscreen

Professional sketch[1]

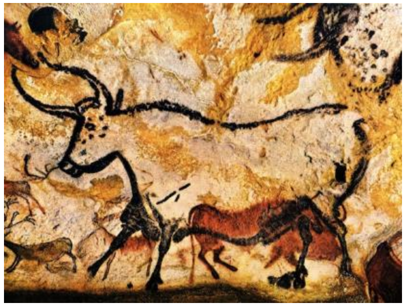


Free-hand sketch

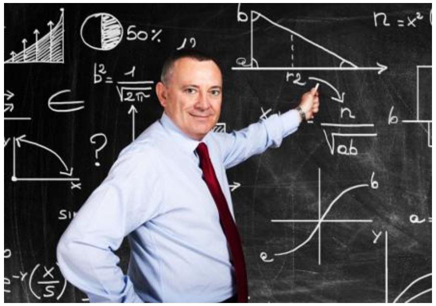
[1] https://s-media-cache-ak0.pinimg.com/236x/9d/0e/a4/9d0ea426351bdfc79043fcd113738def.jpg

Motivation

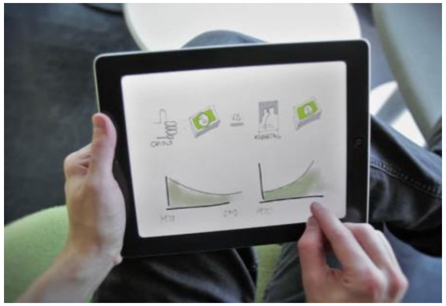
- It has been used as a basic visual communication media since pre-historic times



1500 BC [1]



AD 1800 [2]



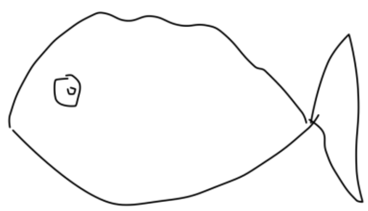
AD 1970 [3]

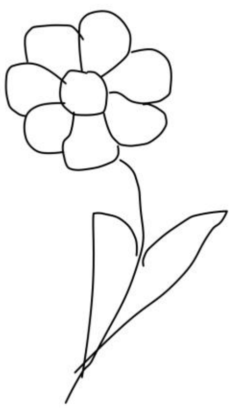
- A wide range of sketch-related applications have been investigated

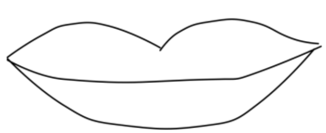
[1] http://blogs.smithsonianmag.com/science/files/2012/12/lascaux-cave-painting.jpg
[2] http://pic.qiantucdn.com/58pic/18/84/12/5655b840744ee_1024.jpg
[3] http://www.arageek.com/wp-content/uploads/Business-sketch-iPad-app.jpg

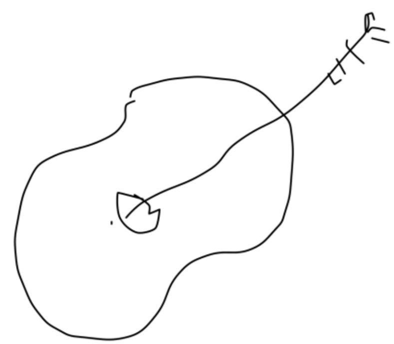
Why difficult?

- Sketches are highly iconic and abstract







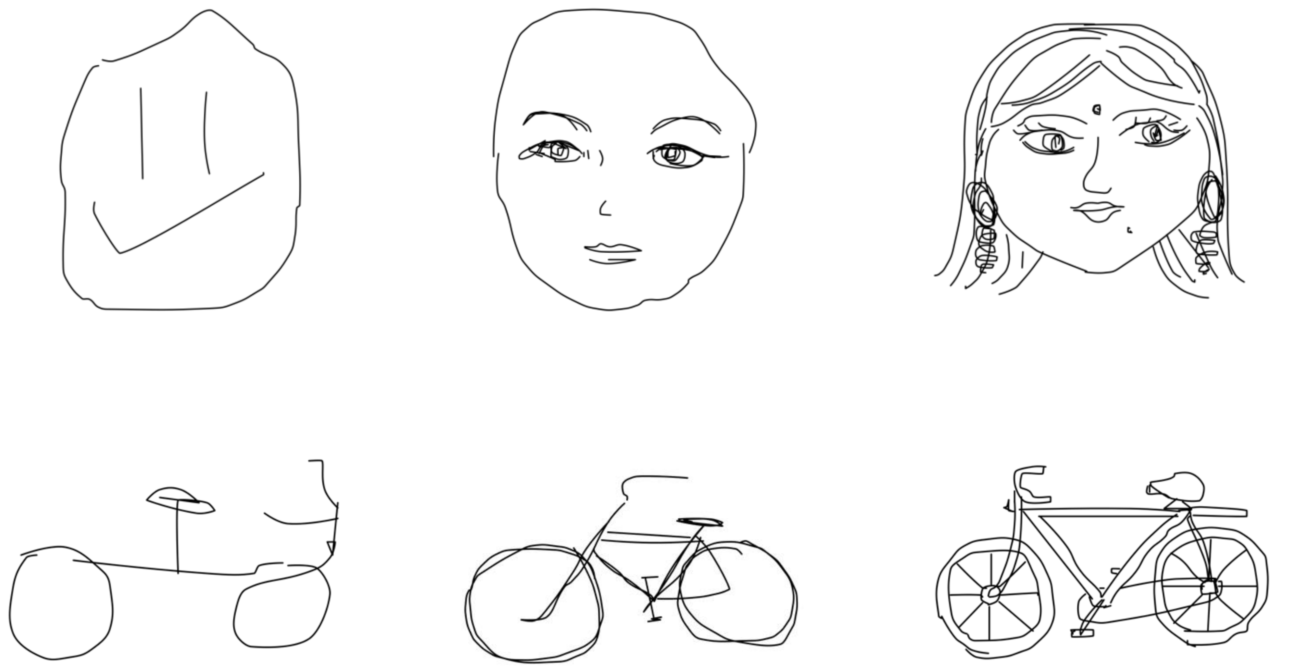


Why difficult?

- Sketches exhibit hugely varied levels of details/abstraction

face

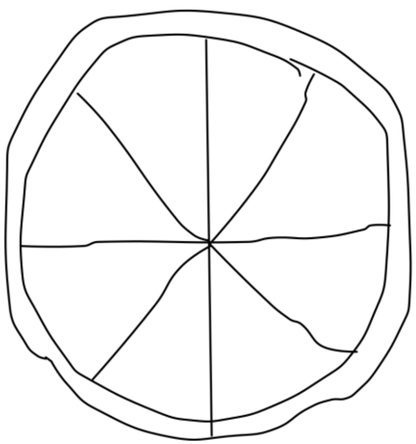
bicycle



Less abstract

Why difficult?

- Sketches lack visual cues, like color, texture etc.



?

Prior Work

Prior work generally follows the conventional image classification paradigm

Feature Extraction → Feature Representation → Classification

Feature Extraction	Feature Representation	Classification
HOG	BoW	SVM
SIFT	FV	…
shape context	star graph	…
…	…	
…	…	

- Limitations
  - No specific or effective features
  - No sequential ordering information

Methodology

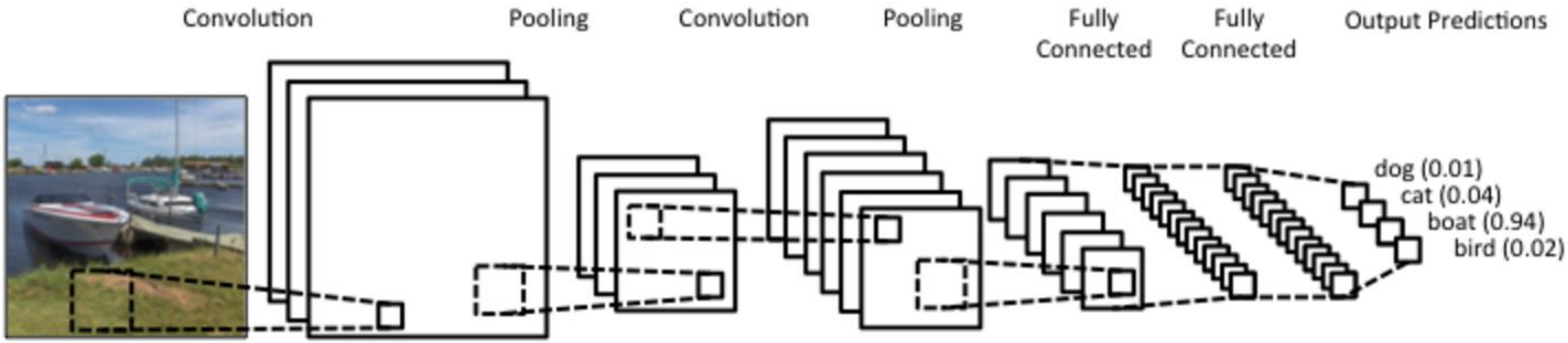
Sketch-a-Net

(1) **Specifically designed CNN** for feature representation learning

(2) **Multi-channel** CNN architecture to embed sequential ordering information

(3) **Multi-scale** network ensemble to address the variability of levels of abstraction

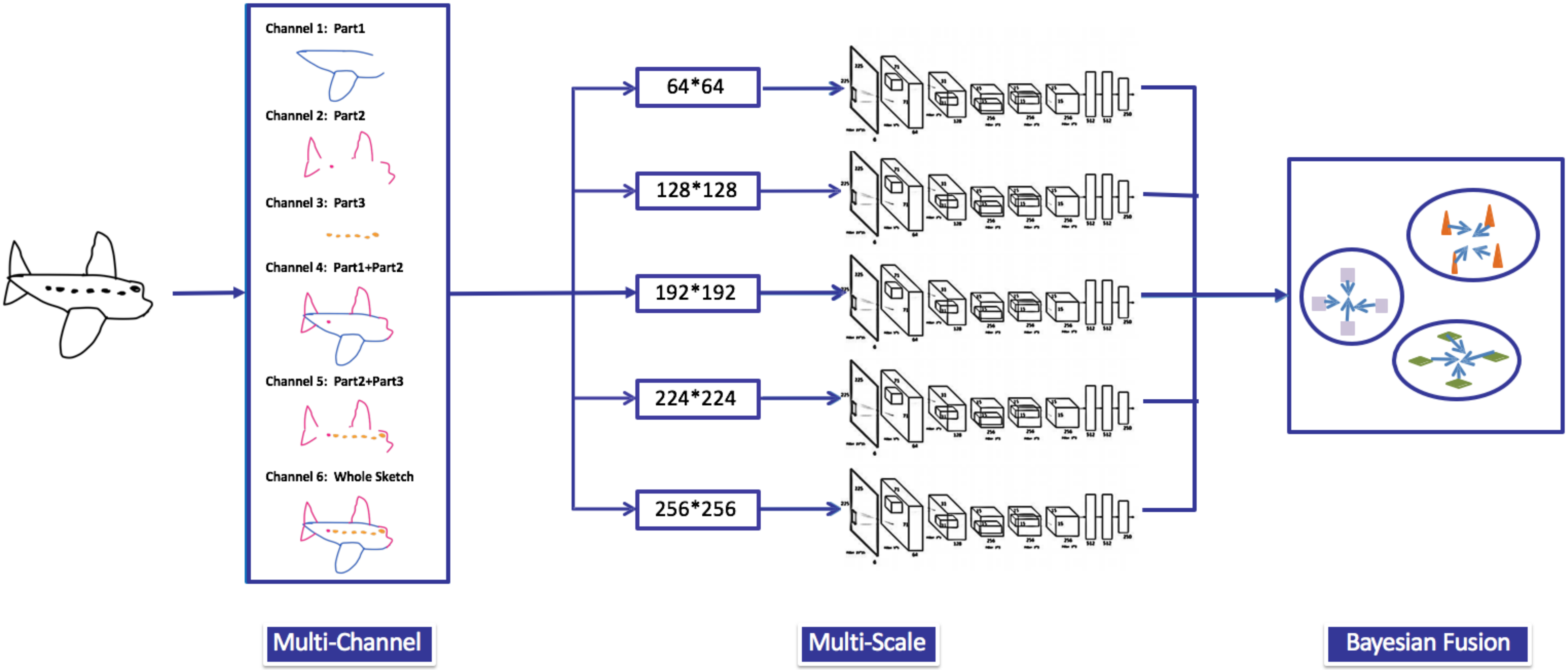
Convolutional neural network (CNN)



Reference: http://d3kbpzbmcynnmx.cloudfront.net/wp-content/uploads/2015/11/Screen-Shot-2015-11-07-at-7.26.20-AM.png

Methodology

Architecture of Sketch-a-Net

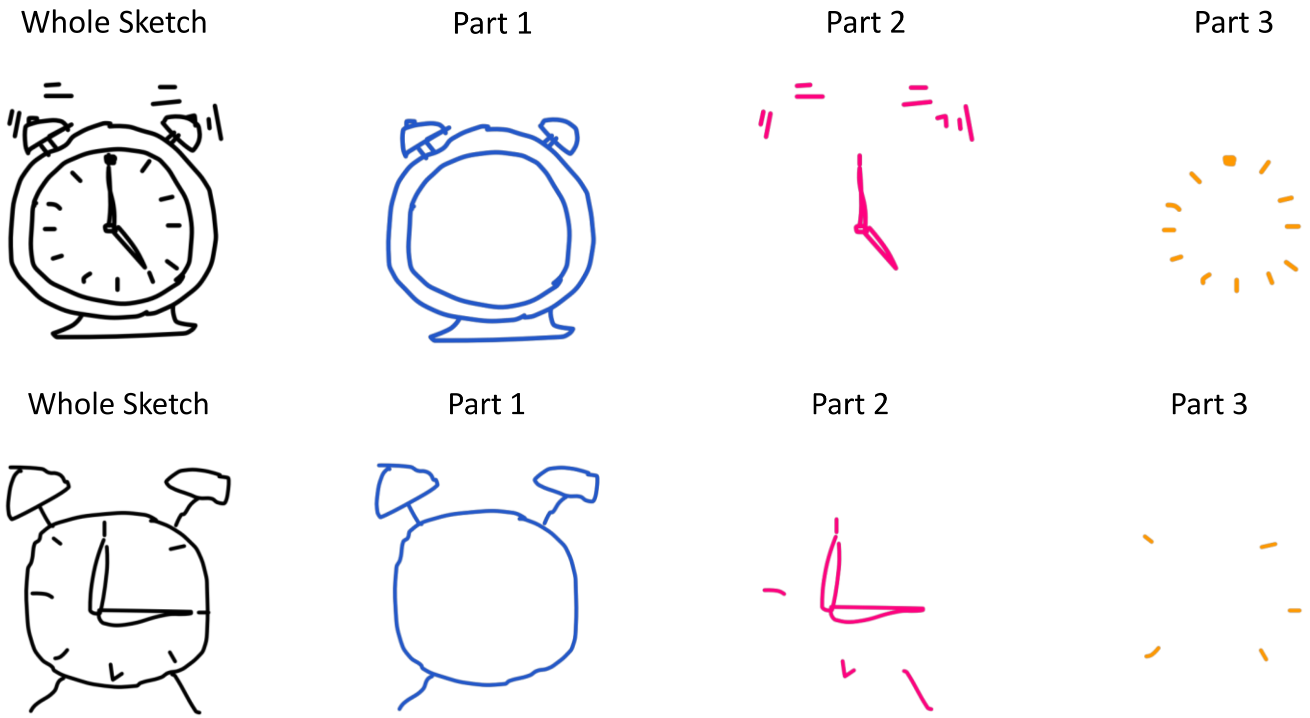


CIS centre for intelligent sensing

Queen Mary University of London

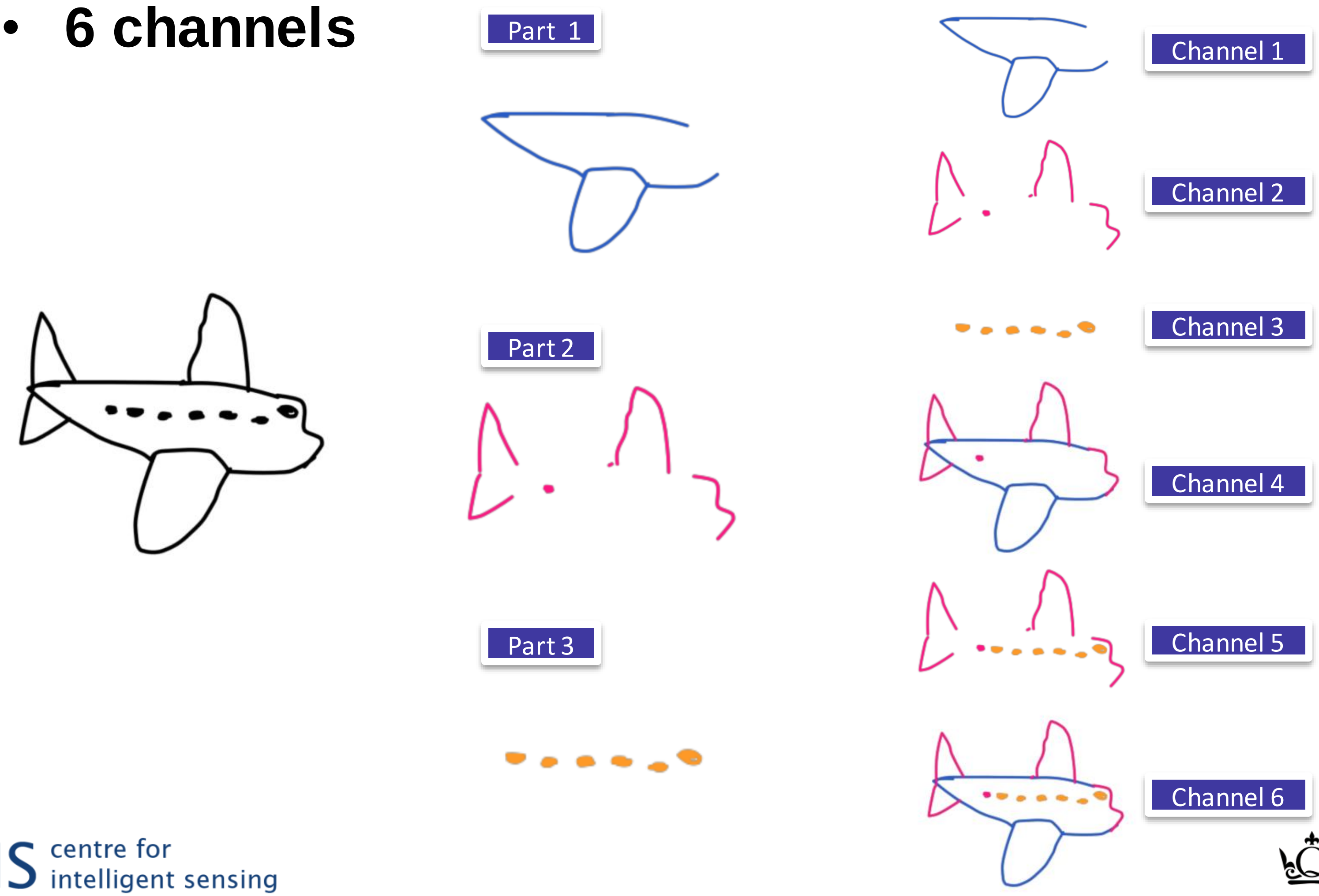
Sketch-a-Net

- Modelling sketch stroke order with multiple channel



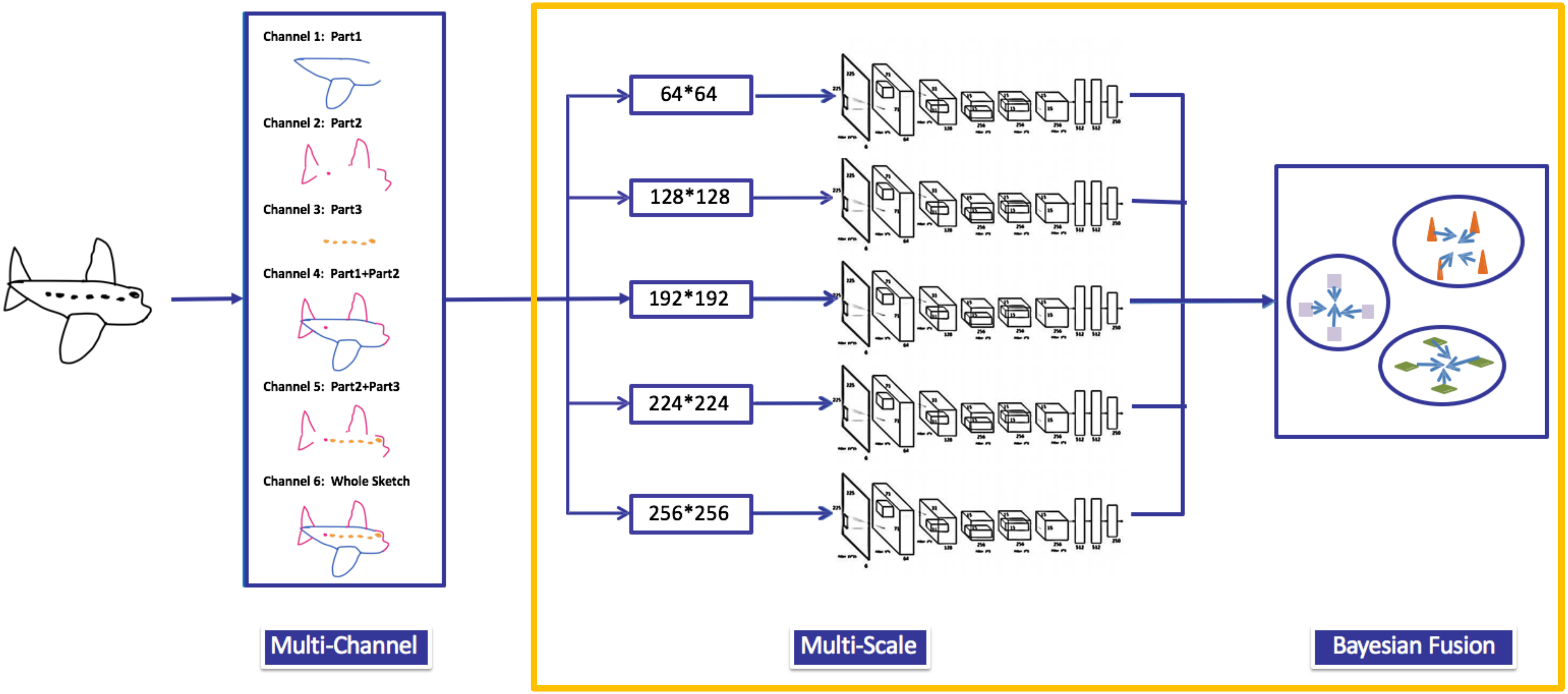
Sketch-a-Net

- Modelling sketch stroke order with multiple channels
- **6 channels**



Sketch-a-Net

- A multi-scale network ensemble with Joint Bayesian fusion



Sketch-a-Net

- How do we get multi-scale network
  - Original size: 256*256
  - Downsample size: 256*256, 224*224, 192*192, 128*128, 64*64
  - Upsample to original size → different blur levels

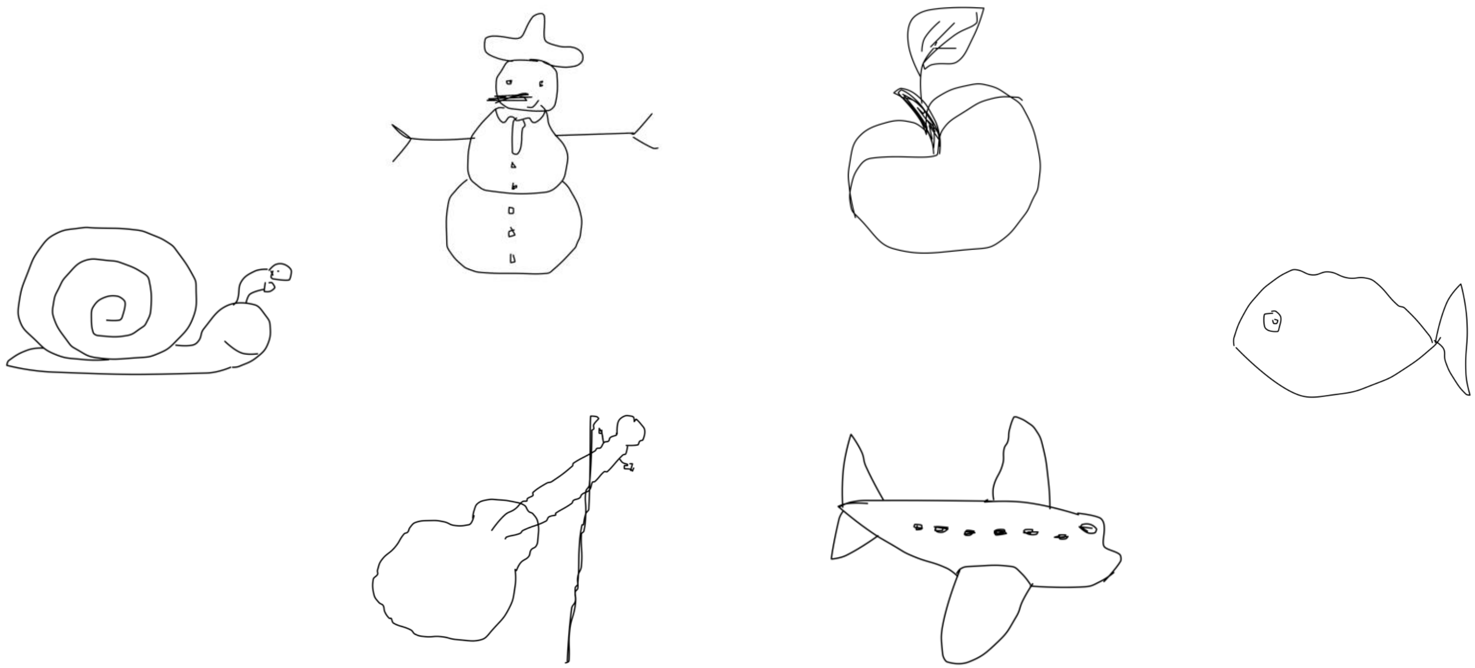
- Joint Bayesian fusion [2]

$$r(x_1, x_2) = \log \frac{P(x_1, x_2 \mid H_I)}{P(x_1, x_2 \mid H_E)}$$

- Testing: using the likelihood ratio as a distance for KNN matching

Experiments and Results

- Dataset: TU-Berlin sketch dataset [3]
  - Collected on Amazon Mechanical Turk (AMT) from 1350 participants
  - 250 categories, 80 sketches per category



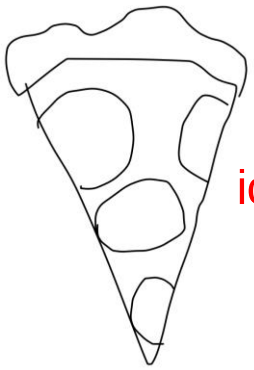
Experiments and Results

Classic	**HOG-SVM [3]**	**Ensemble [4]**	**MKL-SVM [5]**	**FV-SP [6]**	
	56%	61.5%	65.8%	68.9%	
Deep	**AlexNet-SVM [1]**	**AlexNet-Sketch [1]**	**LeNet [7]**	**Sketch-a-Net**	**Human [3]**
	67.1%	68.6%	55.2%	**74.9%**	73.1%

Table 1: Comparison with state of the art results on sketch recognition

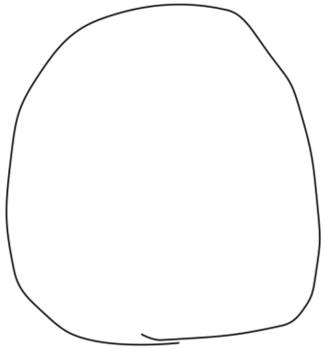
Failure with Bad Drawings

Failure cases

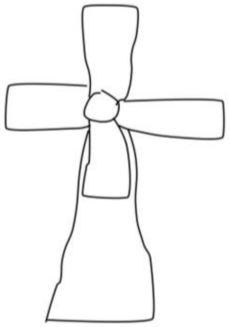


ice-cream-cone
(pizza)

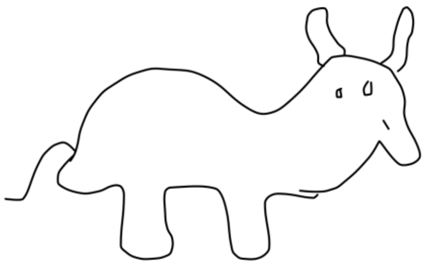
lobster
(bear)



moon
(pizza)



windmill
(Fan)



dog
(pig)

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.

[2] D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun. Bayesian face revisited: A joint formulation. In ECCV, 2012.

[3] M. Eitz, J. Hays, and M. Alexa. How do humans sketch objects? In *SIGGRAPH*, 2012.

[4] Y. Li, Y. Song, and S. Gong. Sketch recognition by ensemble matching of structured features. In *BMVC*, 2013.

[5] Y. Li, T. M. Hospedales, Y. Song, and S. Gong. Free-hand sketch recognition by multi-kernel feature learning. CVIU, 2015.

[6] R. G. Schneider and T. Tuytelaars. Sketch classification and classification-driven analysis using fisher vectors. In *SIGGRAPH Asia*, 2014.

[7] Y. LeCun, L. Bottou, G. B. Orr, and K. Müller. Efficient backprop. *Neural networks: Tricks of the trade*, pages 9–48, 2012.

Questions