Deep Convolutional Neural Network for Computer Vision Products

LI XU, R&D Director
SenseTime Group Limited

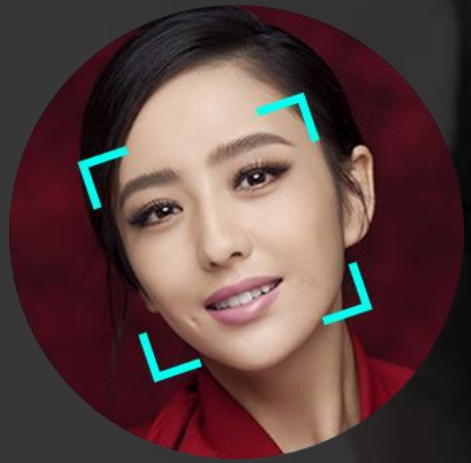
SenseTime Introduction

SenseTime focuses on invention and development of computer vision and deep learning technologies. Our prestige technologies offer sensation and perception being implemented to wide range of system applications, to seize, to analyze and to understand varieties of vision information, as natural as human being & animals.

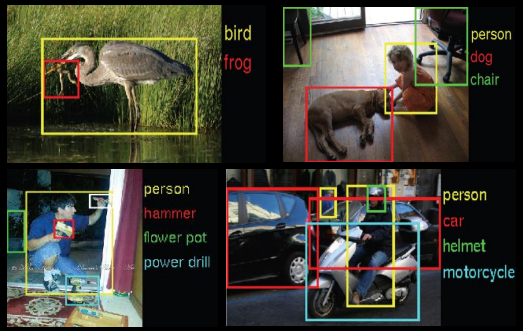
SenseTime is the one of the pioneers in the industries of face recognition, object recognition, image searching, and intelligent monitoring by the virtue of its innovated technologies. By the end of 2014, SenseTime has cooperated with more than 60 well-known organizations in both business and research areas. We were favored by IDG Capital, which is one of the biggest venture capital investor and have successfully closed an investment deal for over millions of dollars.

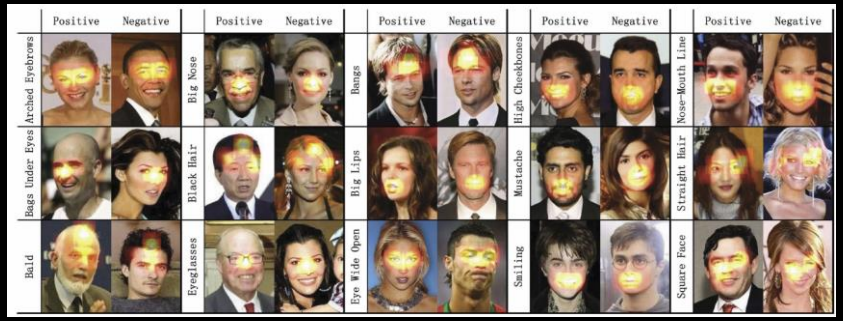
One of the most remarkable breakthrough of SenseTime in 2014 is our core technology - face recognition, has now been developed to, and reached over 99% accuracy rate, and that figure shows it performs even better than natural human's recognition.

SENSETIME



SENSETIME





Big Visual Data

NVIDIA GPUs

Deep Learning

Big Visual Data

Our Awards	Conference	Best Paper
Machine Learning	NIPS ’10	Best Student Paper
Computer Vision	CVPR’09	Best Paper
Artificial Intelligence	AAAI’ 15	Best Student Paper

SENSETIME

NVIDIA GPUs

NVIDIA

- 2GPUs → 300 GPUs
- CVPR: 14/29 deep learning papers published in the whole world. (12'-14')

Deep Learning

Detection

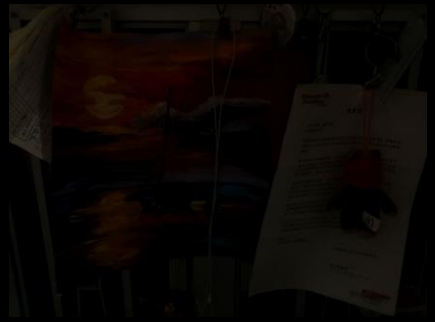
- Pedestrian detection
- Human pose estimation
- Facial keypoint detection

Segmentation

- Face parsing
- Pedestrian parsing

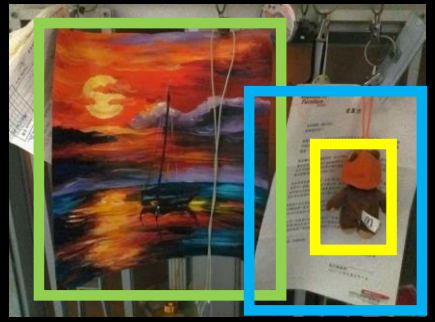
Recognition

- Face attribute recognition
- Human identity recognition across camera views



Capturing

Enhancement



Localization

Oil Painting

Paper

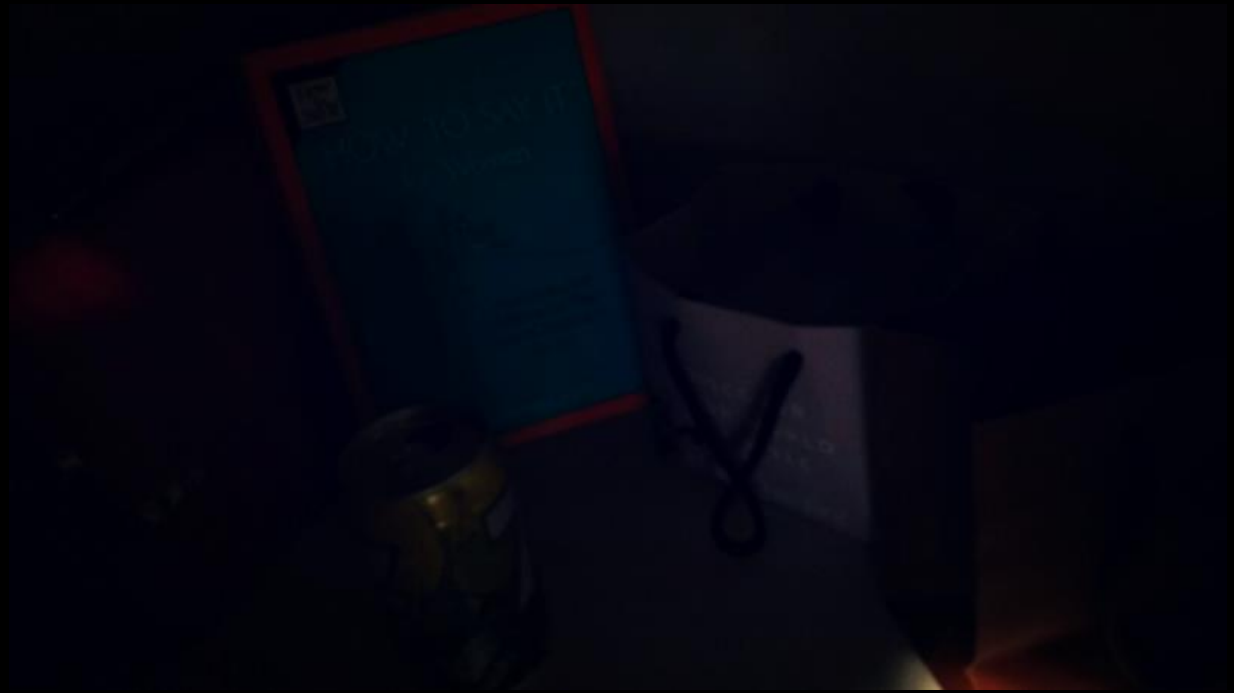
Toy

Classification

SEEING

UNDERSTANDING

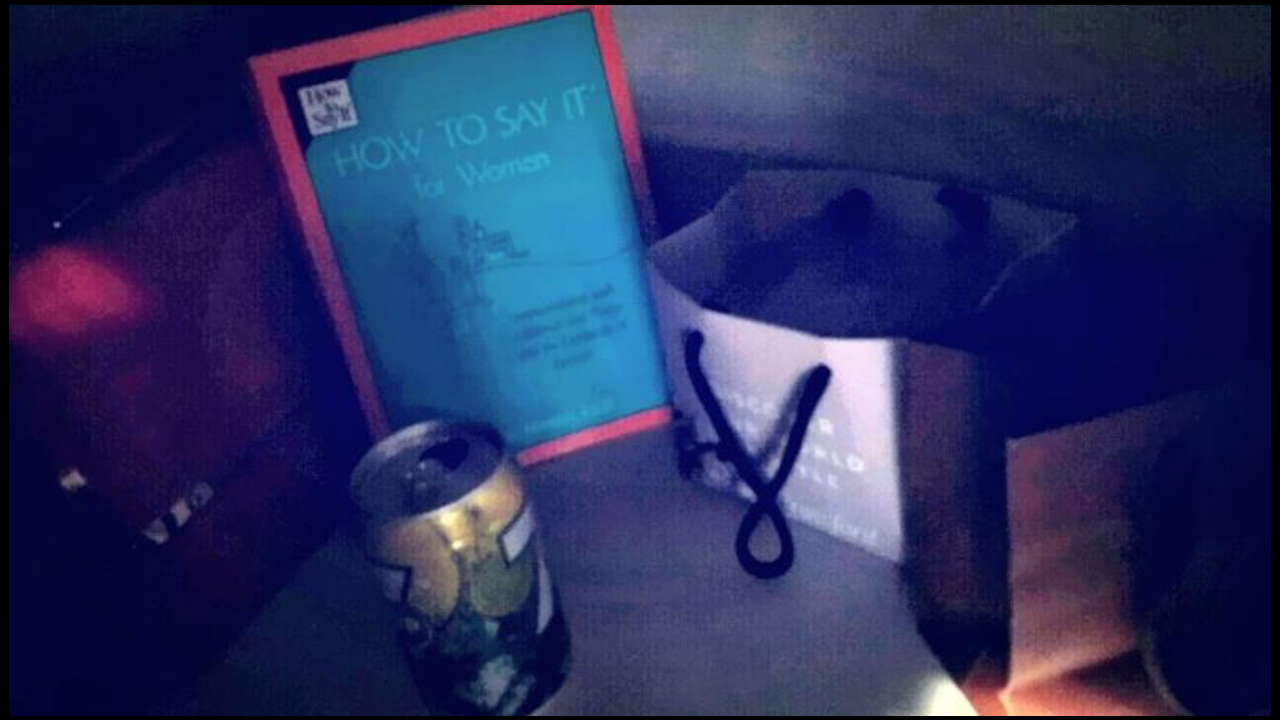
Seeing is Believing



- Face
- Book
- Bag

The Photo is Captured by an Android Phone with Baidu SuperCamera

Seeing is Believing



- A Book

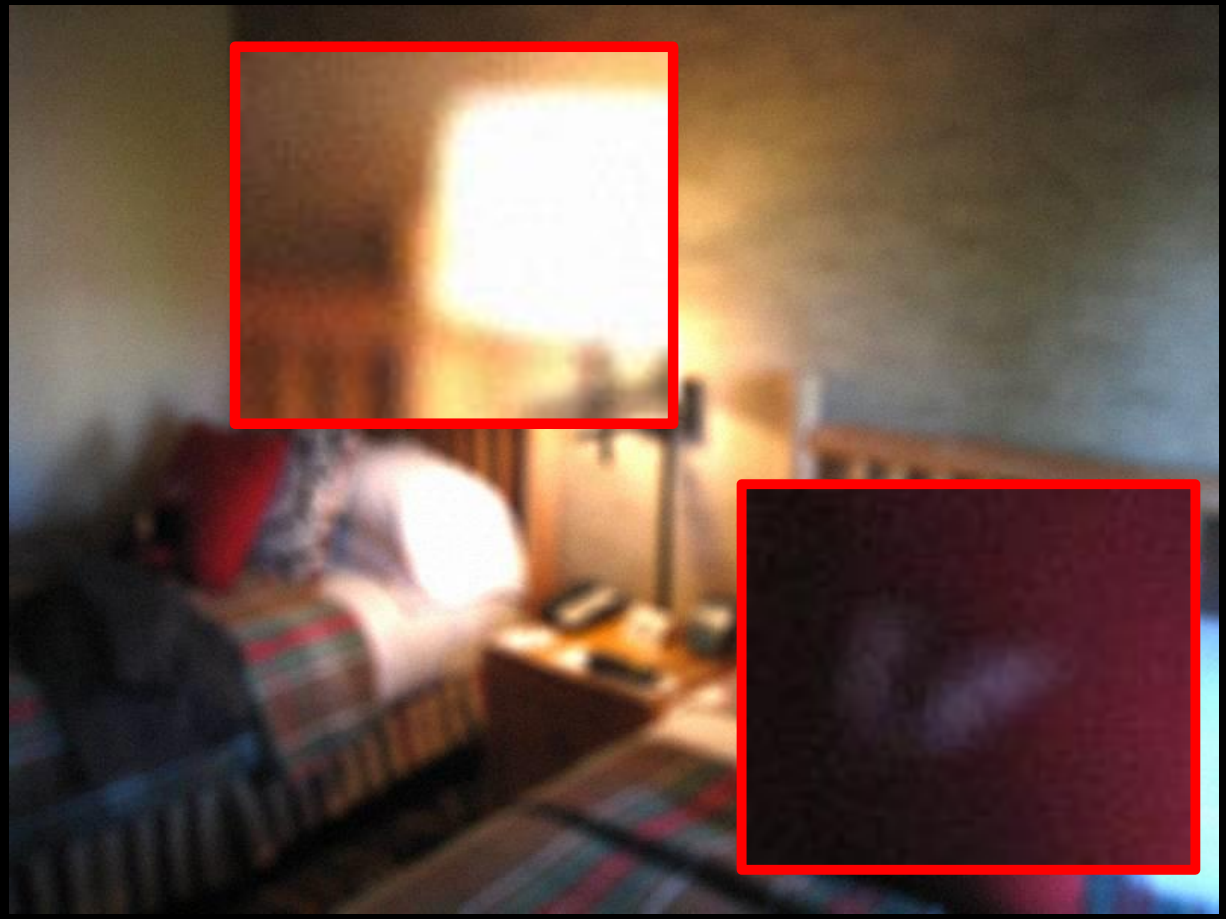
  “How to say it for woman”
- Paper Bags
- 7-UP

The Photo is Captured by an Android Phone with Baidu SuperCamera

Seeing is Believing

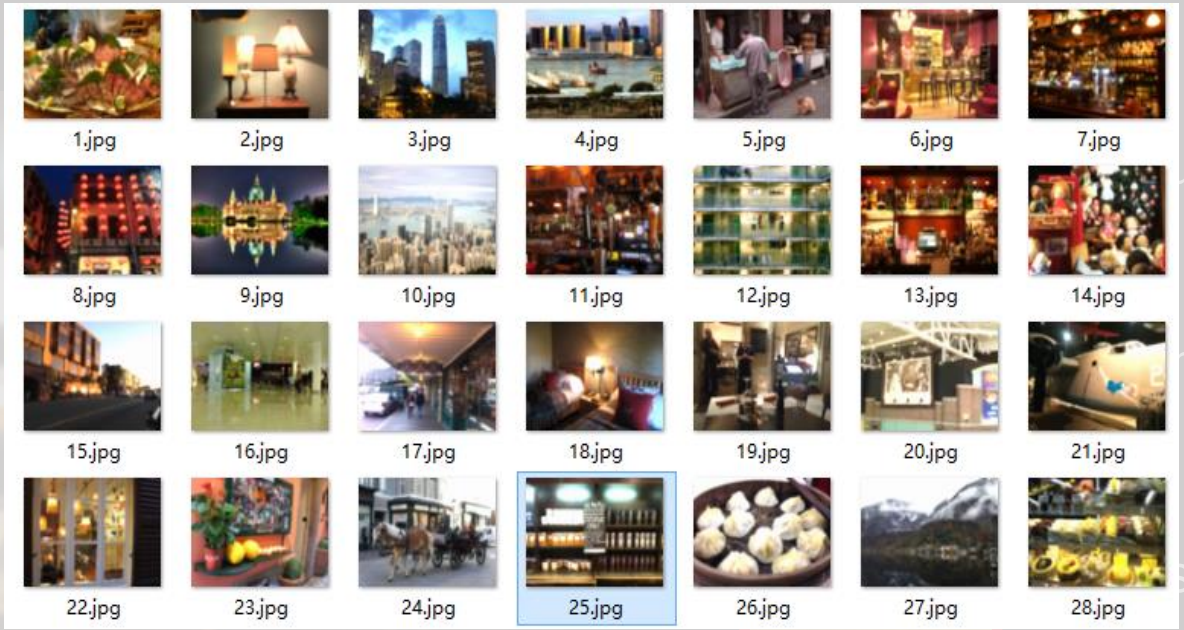
Seeing is Believing

SENSETIME



DCNN for Low-Level Vision

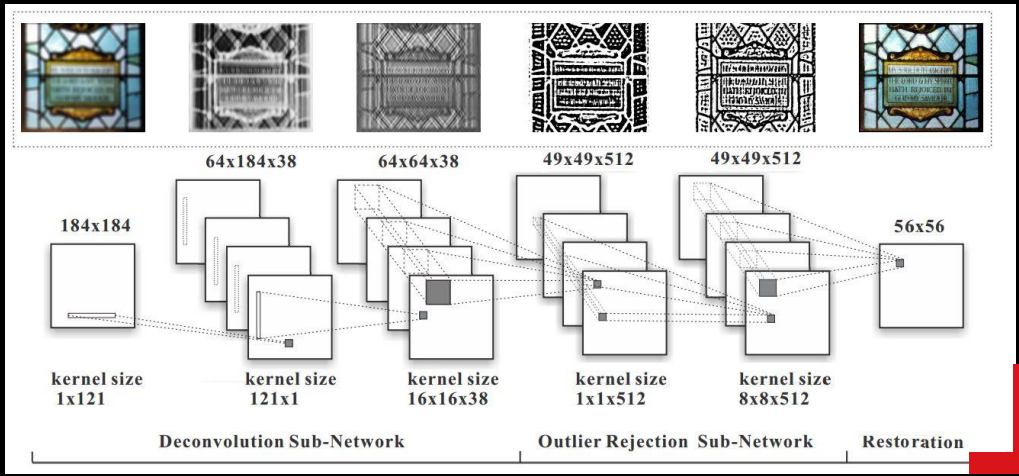
- Data: Big data with real-world degradation



DCNN for Low-Level Vision

- Data: Big data with real-world degradation
- Architecture: use domain-specific knowledge

  A Large Kernel Deep CNN for deconvolution



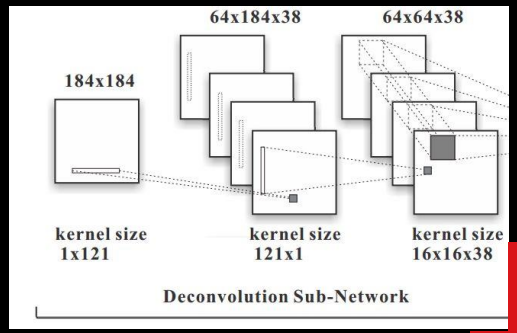
-121x121 spatial support based on kernel SVD

DCNN for Low-Level Vision

- Data: Big data with real-world degradation
- Architecture: use domain-specific knowledge
- Training: Better initialization, GPU acceleration

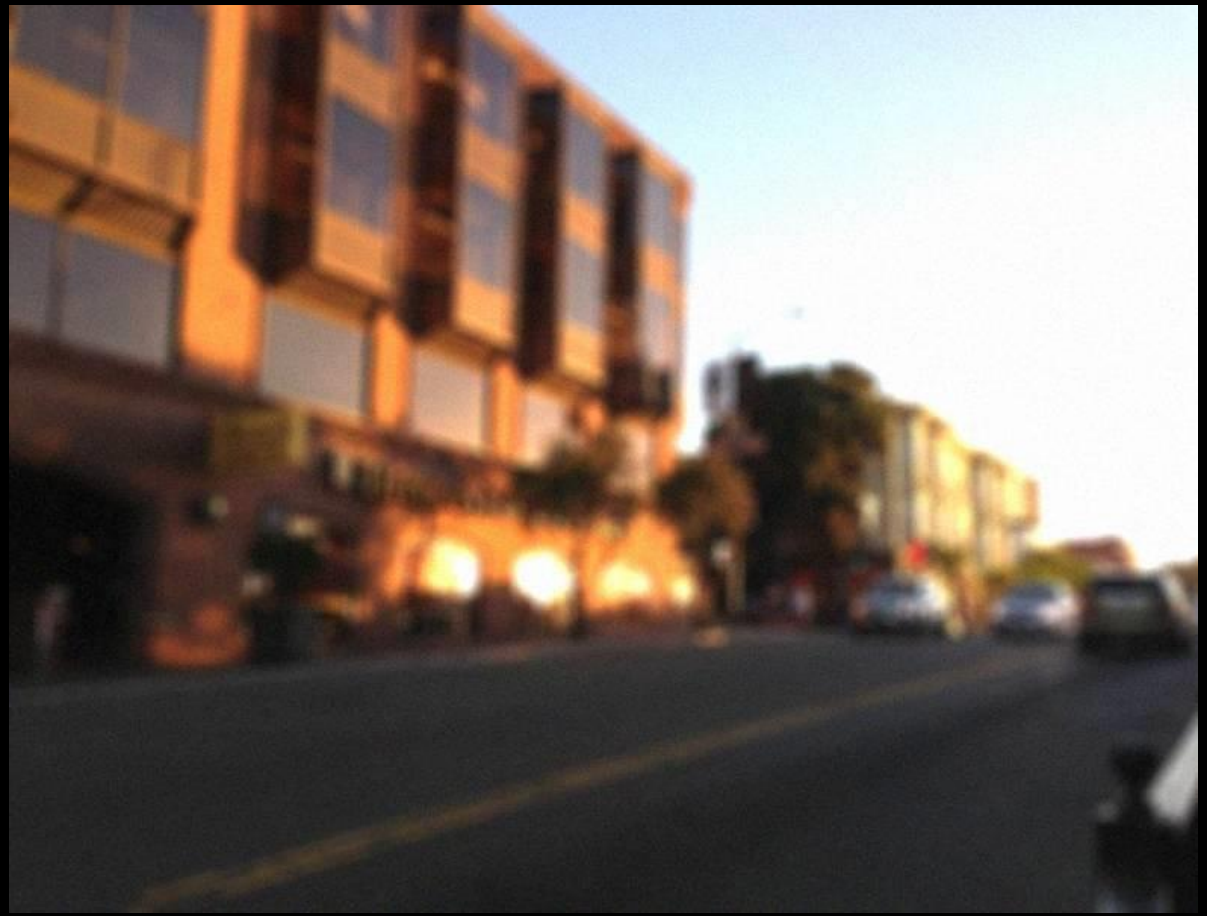
12-20 hours

NVIDIA



A novel weights initialization
Supervised pre-training

SENSETIME



SENSETIME

图像服务

ImagePro

智能对比增强

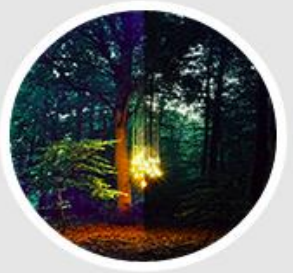
Smart Contrast

照片快速去噪

Fast Denoise

图像前景分割

Grabcut

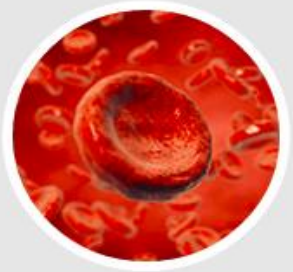


自动提高亮度

Night Enhance

超分辨率

Super Resolution Upscale



照片细节增强

Detail Enhance

图像智能补全

Inpainting

简单色彩转移

Naive Color Transfer

Understanding: Localization & Classification

ImageNet Large Scale Visual Recognition Challenge 2014

Ordered by mean average precision

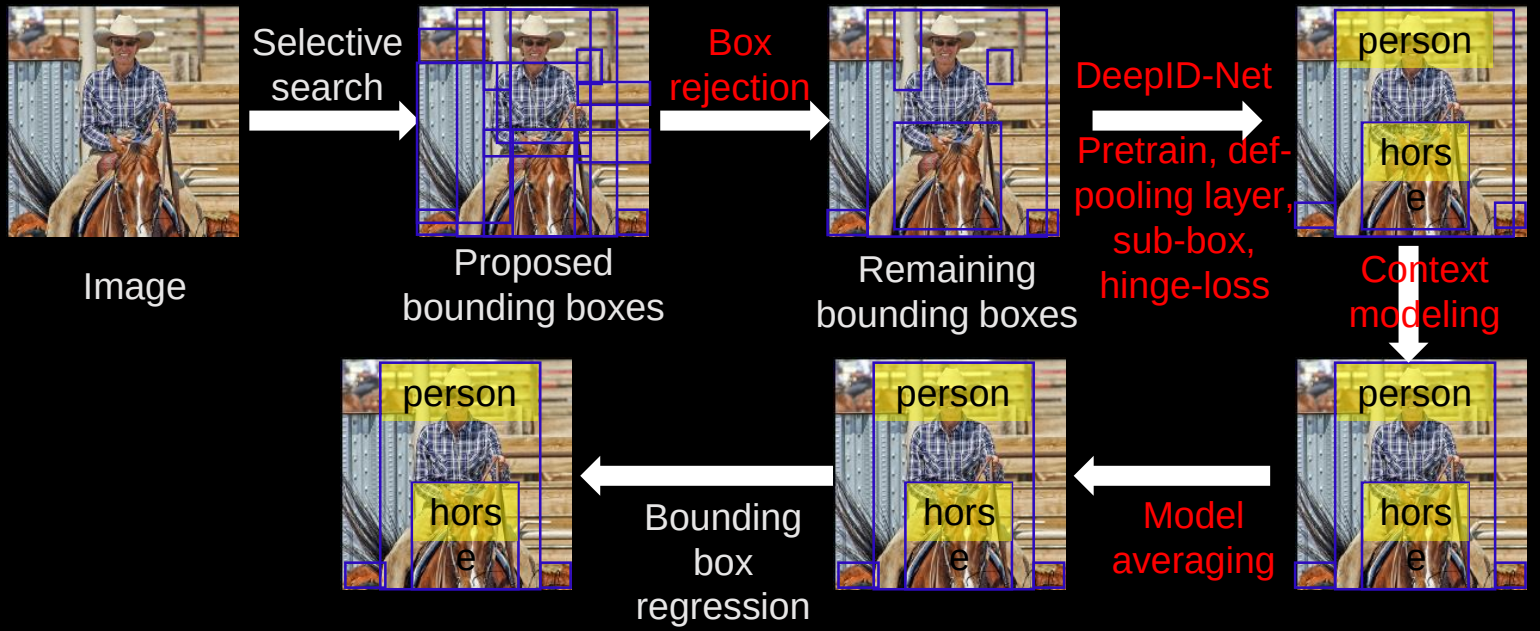
Team name	Entry description	Description of outside data used	**mean AP**	Number of object categories won
GoogLeNet	Ensemble of detection models. Validation is 44.5% mAP	Pretraining on ILSVRC12 classification data.	0.439329	142
CUHK DeepID-Net	*Combine multiple models described in the abstract without contextual modeling. The training data includes the validation dataset 2.*	*ImageNet classification and localization data*	*0.406998*	---
CUHK DeepID-Net	Combine multiple models described in the abstract without contextual modeling	ImageNet classification and localization data	0.406659	29
Deep Insight	Combination of three detection models	Three CNNs from classification task are used for initialization.	0.404517	27
CUHK DeepID-Net2	*Combine multiple models described in the abstract without contextual modeling. The training data includes the validation dataset 2.*	*ImageNet classification and localization data*	*0.40352*	---
CUHK DeepID-Net2	*Combine multiple models described in the abstract without contextual modeling*	*ImageNet classification and localization data*	*0.403417*	---

DCNN for Object Recognition

- A Novel Data Generation for Pre-training

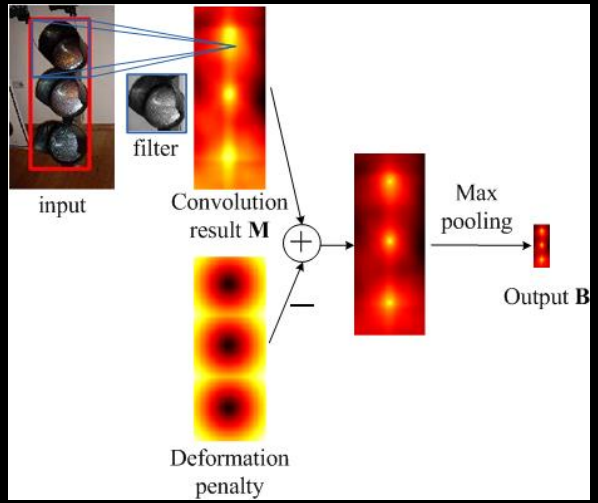
DCNN for Object Recognition

- A novel DCNN pipeline



DCNN for Object Recognition

- A deformable constraint pooling



DCNN for ImageNet

- Training
- 4-core 3.3G CPU
  - 70 seconds /image
  - 50 months for training
- Titan GPU
  - 1s / image
  - 21 days for training

Face Verification

- #1 on LFW, with mean accuracy ~99.53%
- Human Performance on LFW ~ 97.53%

Nicole Kidman | Nicole Kidman

Coo d'Este | Melina Kanakaredes

Jim O'Brien | Jim O'Brien

LFW Ranking

Methods	Accuracy
FR+FCN	0.9645 ± 0.0025
DeepFace-ensemble	0.9735 ± 0.0025
DeepID	0.9745 ± 0.0026
GaussianFace	0.9852 ± 0.0066
DeepID2	0.9915 ± 0.0013
DeepID2+	0.9947 ± 0.0012
DeepID3	0.9953 ± 0.0010

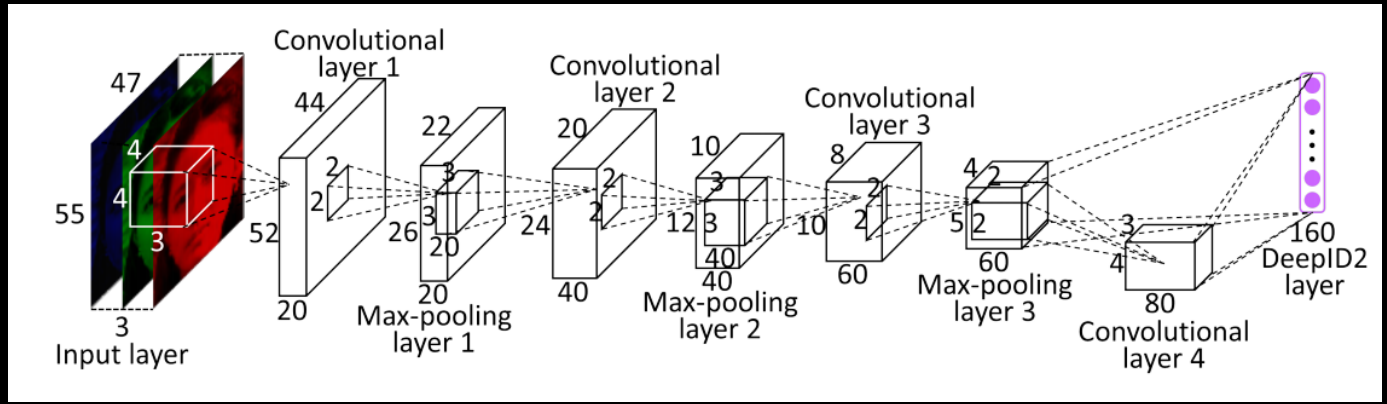
DCNN for Face Recognition/Verification

- 10,000+ Class

  Better generalization for verification

- Joint Identification-Verification

  Reduce intra-person variation

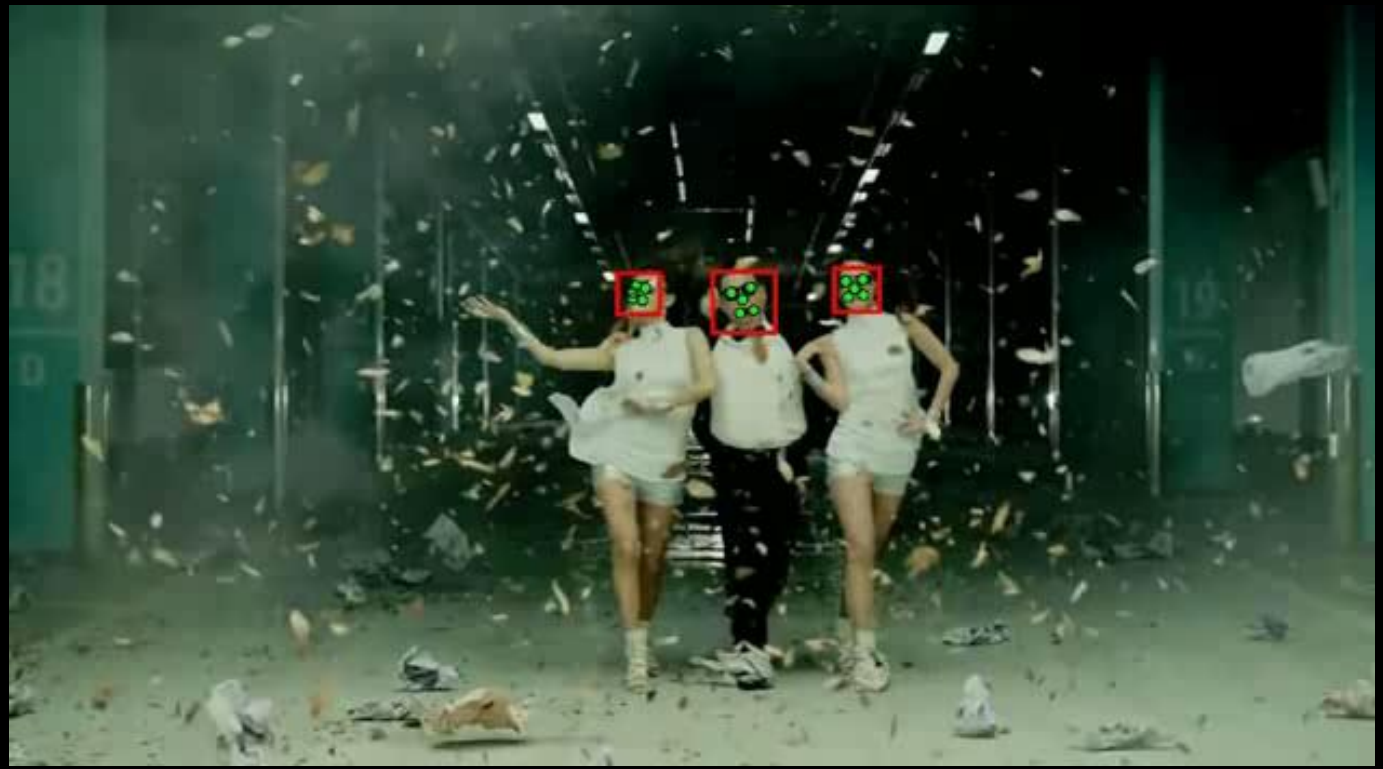


DCNN for Face Recognition/Verification

- Learning by predicting 10,000+ Class
- Joint Identification-Verification
- Over-complete representation

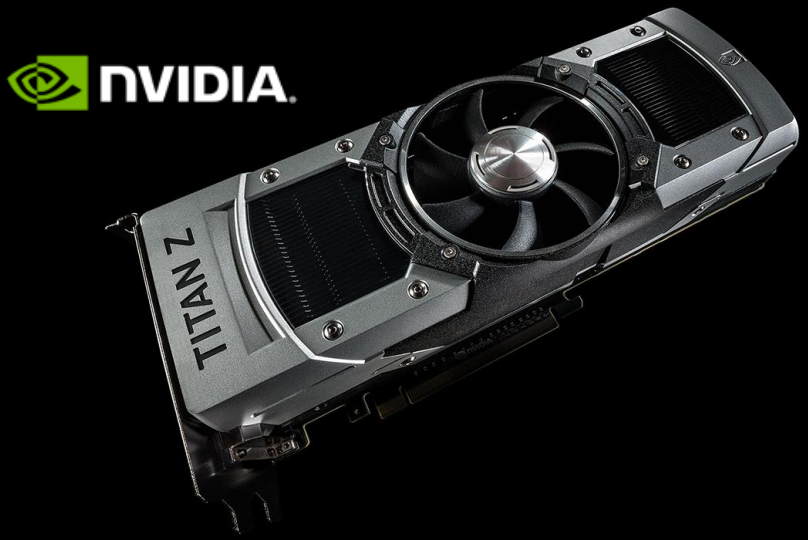
  Learning features from multiple cropped face regions

Robust Face Detection



DCNN for Face Recognition/Verification

- CPU cores @2.66GHz: ~20 days
- Titan Z GPU: 6 hours



SENSETIME

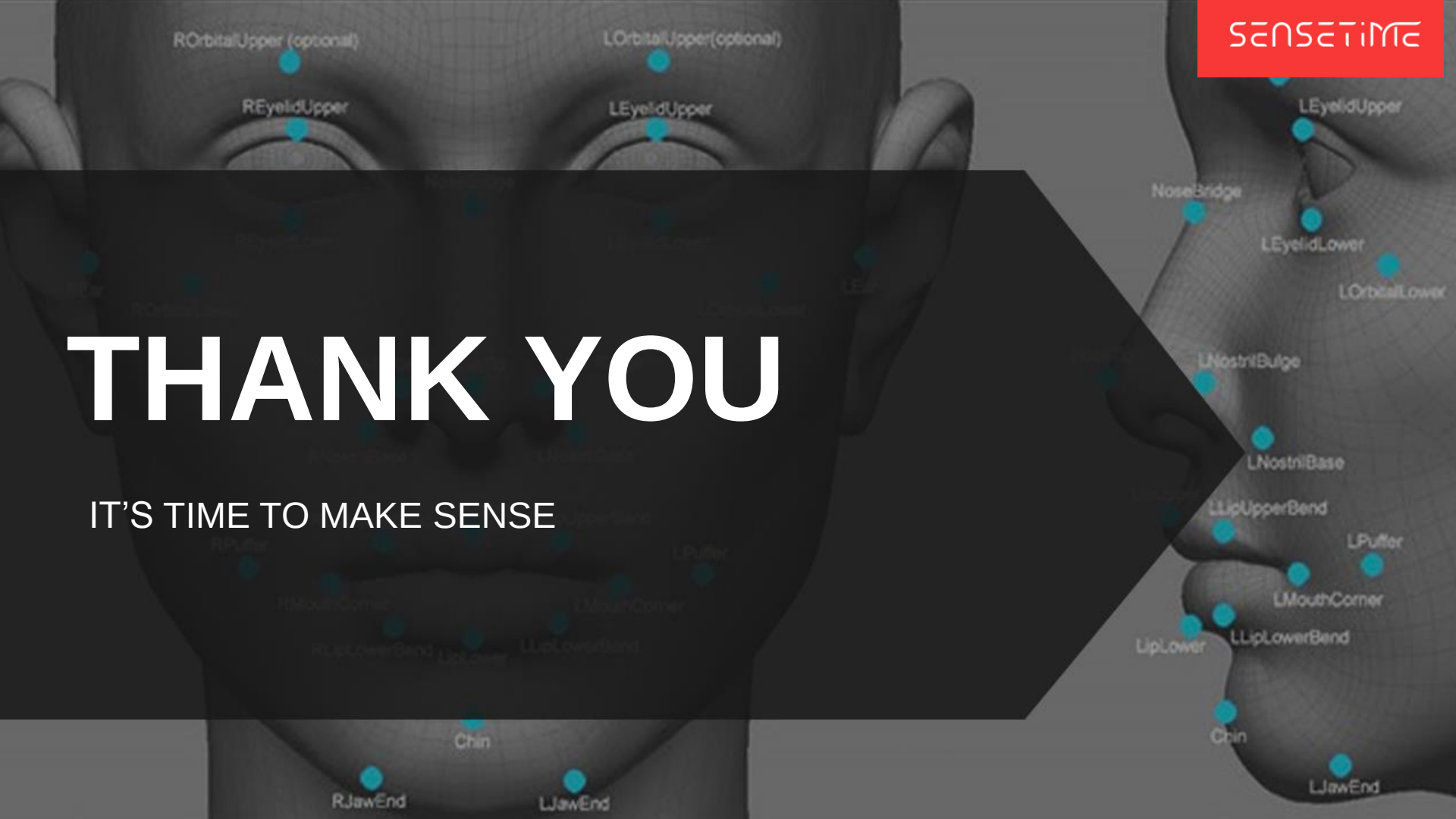
Computer Vision Solutions

SEEING

- Low-light Enhancement, Visibility Enhancement (haze, dust) , Super Resolution, Blur Removal

UNDERSTANDING

- Face detection, recognition, verification, Object Recognition, Gesture recognition, Pedestrian Detection, Crowd Analysis



SENSETIME

THANK YOU

IT’S TIME TO MAKE SENSE