Deep Vision for *Human Vision*

Yu Qiao[1] and Zhe Wang[2]

1. Shenzhen Institutes of Advanced Technologies, Chinese Academy of Sciences
2. Sensetime Limited

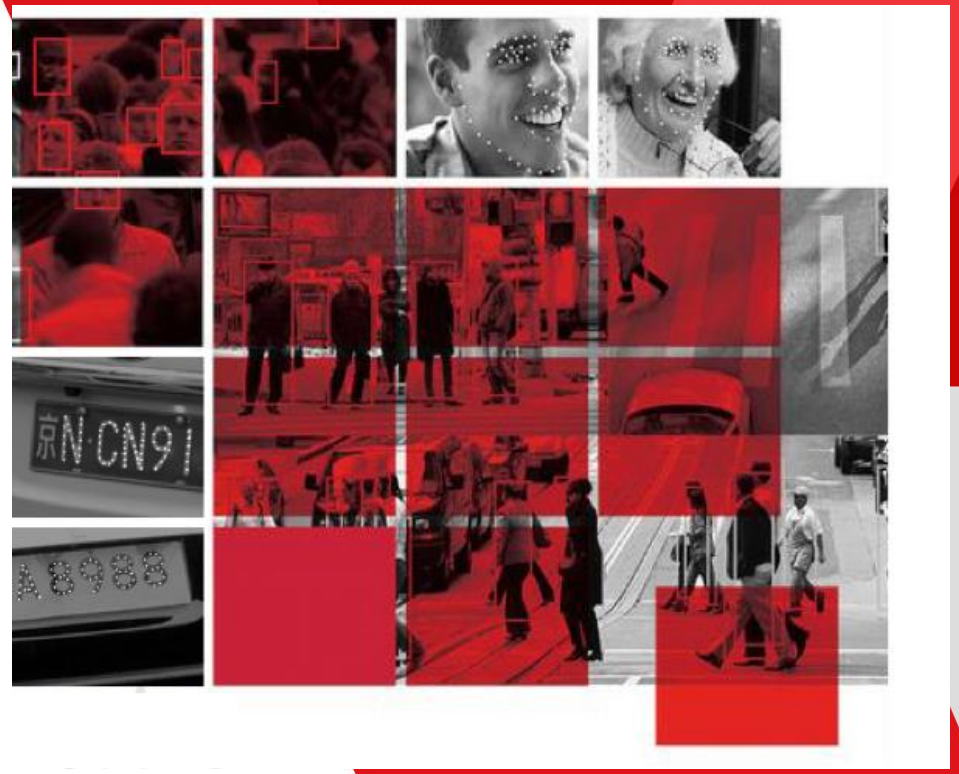
SENSETIME
商 汤 科 技

SenseTime: Who We Are

SenseTime is an AI company focusing on:

① Computer Vision

② Deep Learning



SenseTime developed a proprietary deep-learning platform, the only one in China

SENSETIME
商汤科技

SenseTime – The Largest AI Unicorn in China

Company Profile

Biggest AI Unicorn
Most Funded & Valued Over $1.5B

The ONLY
Proprietary Deep Learning Platform in China

20 years
Research Experience of Founding Team

500+ Employees
120+ PhDs, the Biggest AI R&D Team in Asia

8 Offices
5 in mainland, 1 in HK, 2 in Japan

Performance

Revenue
Industry No.1

Largest
Chinese Algorithm Provider

Leader
In Various Vertical Industry

400+
Major Clients

400 Million
People Used Our FaceRec Tech

Leading Technologies


Facial Recognition


Video Image Analysis

Augmented Reality


Text Recognition


Autonomous Driving


Medical Image

SENSETIME 商汤科技

Leader in R&D, Talent, and Business

Research Achievements

- Core team has 20+ years of research experience. One joint lab was named 2016 Top 10 AI Lab in the World
- 400+ research papers published at top conferences
- Beat Facebook to be the first team that achieved better accuracy than human eye with facial recognition AI
- Competed with Google at ImageNet Large Scale Visual Recognition Challenge

CVPR 2016 kaggle IMAGENET

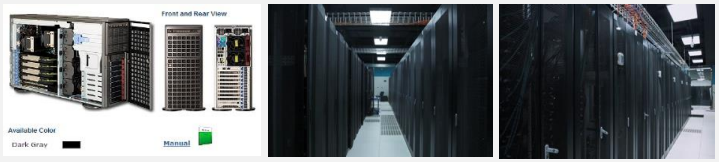
Deep Talent Pool

- The largest deep learning research team in Asia
- 18 professors, 120+ PhDs, and 300+ masters/bachelors from top universities in the world
- One of the largest AI research teams in the world

MIT Stanford Tsinghua CUHK SIAT

The Only Deep Learning Platform in China

- Developed deep learning platform including both software framework and supercomputing system
- Significantly reduced the R&D cost and shortened the time needed in training highly sophisticated models

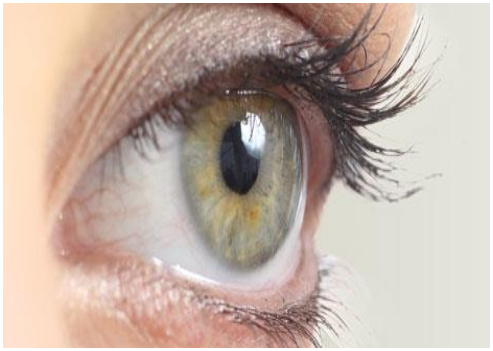


#1 in AI Commercialization

SenseTime's 2016 revenue topped the industry. Customers:

- Video surveillance providers
- Banks and financial institutions
- Smart phone OEMs
- Mobile application developers
- Internet service providers
- Robot makers
- Government agencies & public security authorities

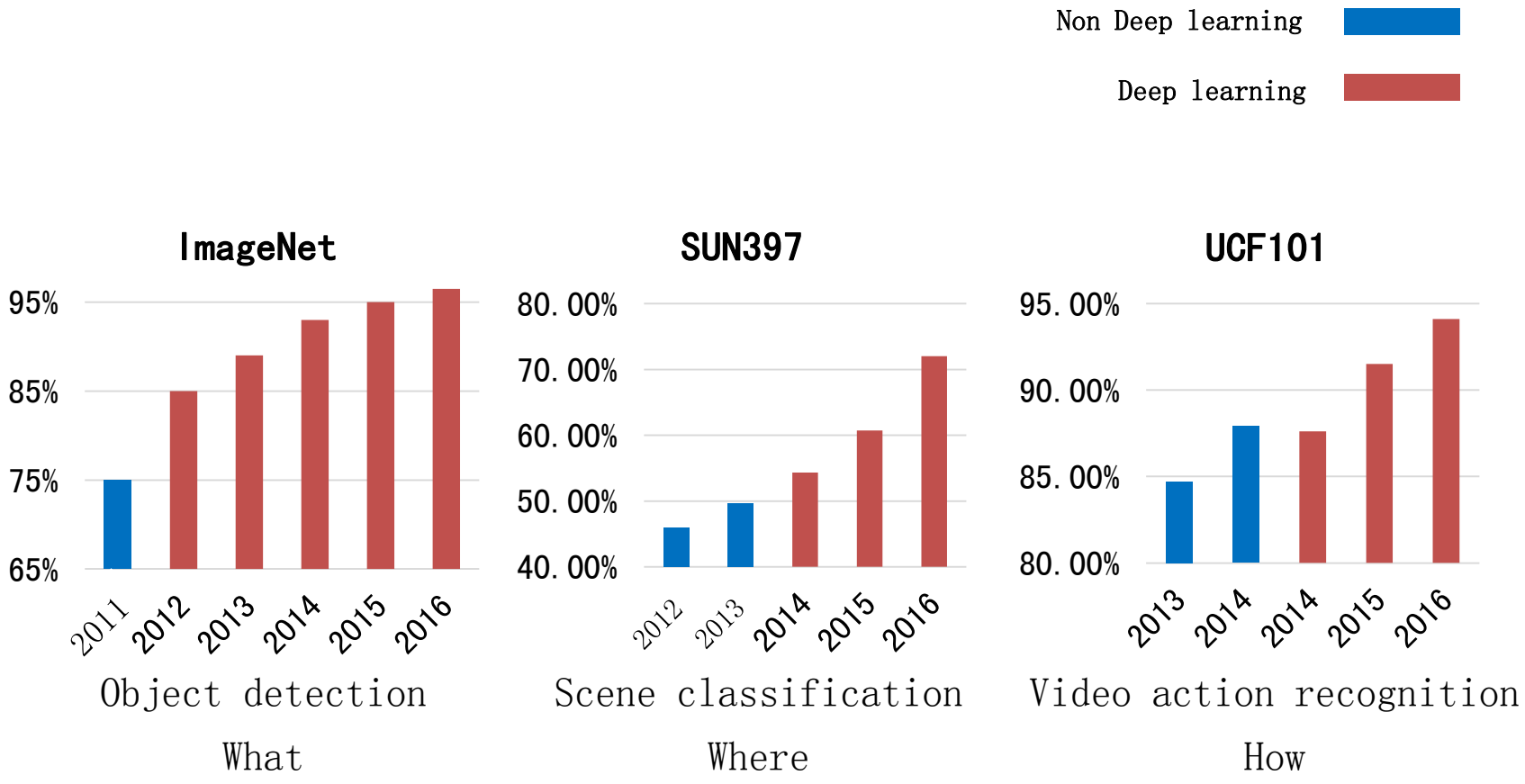
Computer Vision v.s. Human Vision



V. S.

Human vision		Computer vision
• Evolution result	·····	• A sub area of AI
• (~520m yrs)	·····	• (~60 yrs)
• Bio-chemical events	·····	• Electrical computation
• Powerful, robust and reliable	·····	• Still needs requirements

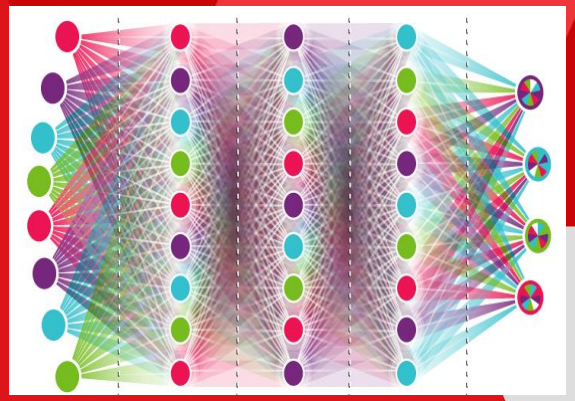
Computer vision vs human vision



Three basic problems of computer vision

Strengths of deep learning

Big data + Powerful computation + Deep models



Human vision

Limited training data + Limited computation + Biology constraints

SENSETIME
商汤科技

Surpass human performance

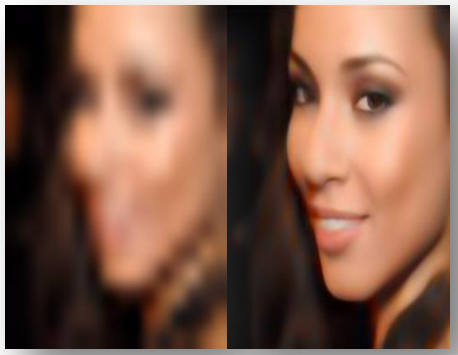
ImageNet: Object Classification
DL 97.21%: Human 94.9%

LFW: Face Verification
DL 99.6%: Human 97.3%

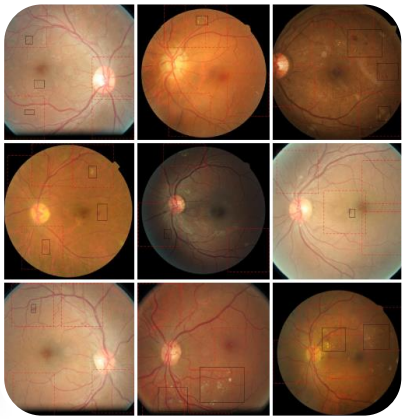
SENSETIME
商汤科技

Can deep learning help, enhance or assistant human vision?

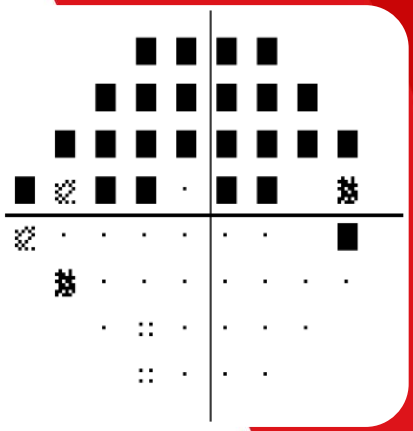
Deep Learning for Human Vision



Super-resolution &
Image enhancement



Diabetic
Retinopathy
Diagnosis



Glaucoma
Diagnosis

Skin lesion
analysis

SENSETIME
商汤科技

Super-resolution and Image Enhancement

SENSETIME
商 汤 科 技

Image enhancement for low-light photos

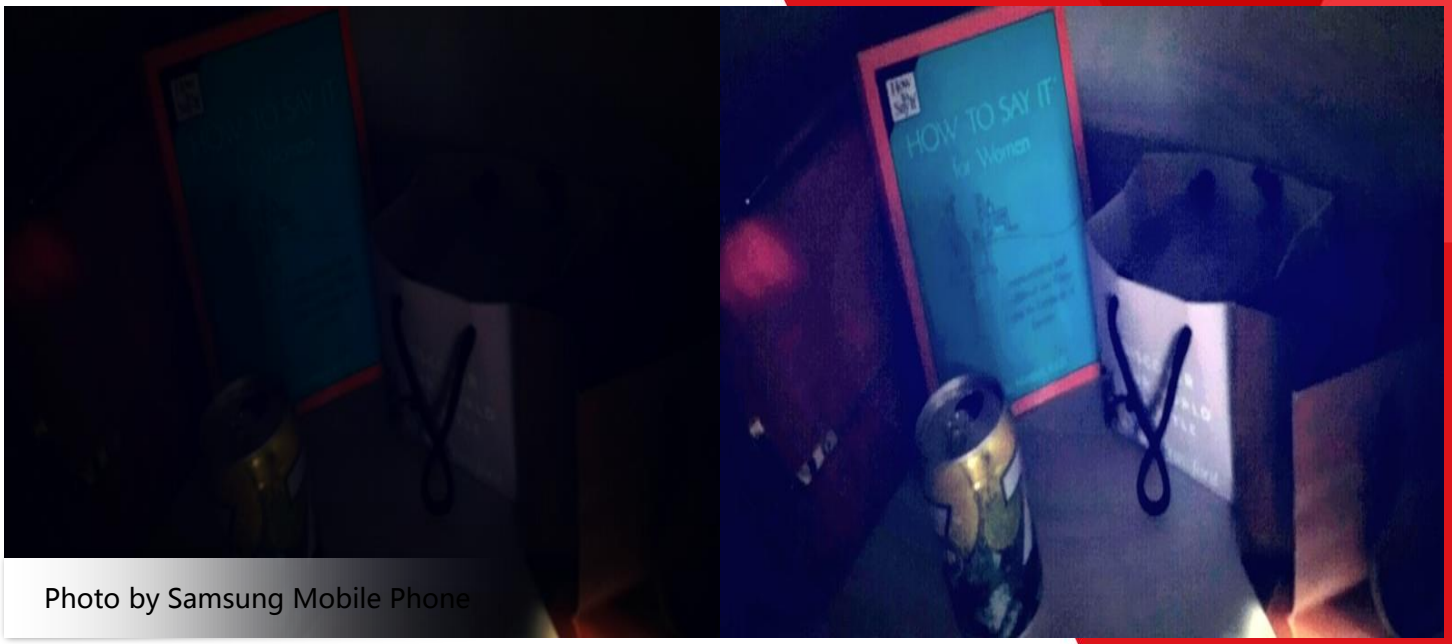
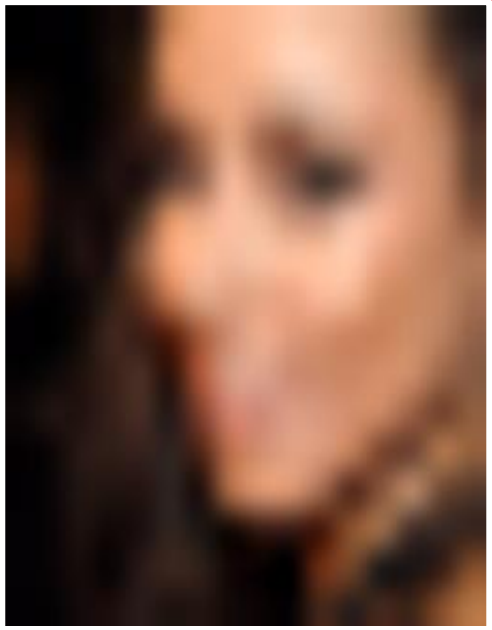


Photo by Samsung Mobile Phone

SENSETIME
商 汤 科 技

Image Super-resolution



SENSETIME
商汤科技

Image deblurring

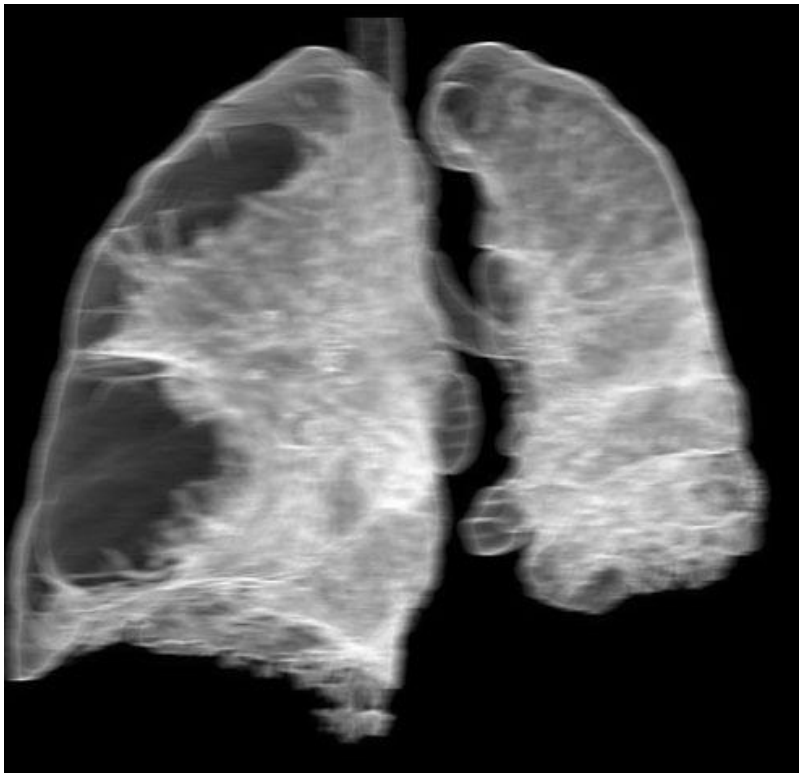
Vieux-Port @ Marseille, 2008 by Nikon D70

SENSETIME
商 汤 科 技

Deblurring for medical image

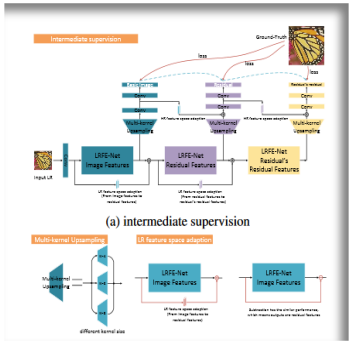
SENSETIME
商汤科技

Deblurring for medical image

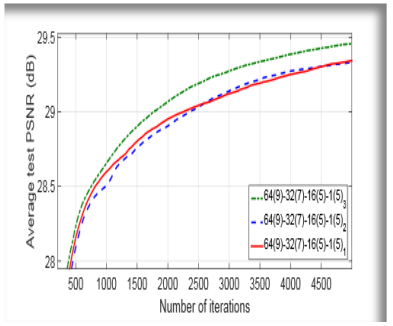


Paper and Challenge Results

- “Deep Convolutional Neural Network for Image Deconvolution “, NIPS 2014
- “On Vectorization of Deep Convolutional Neural Networks for Vision Tasks”, AAAI 2015
- “Deep Edge-Aware Filters”, ICML 2015
- “Shepard Convolutional Neural Networks”, NIPS 2015
- “Learning a Deep Convolutional Network for Image Super-Resolution”, ECCV 2014 (539 citations)
- “Image Super-Resolution Using Deep Convolutional Networks”, TPAMI 2015 (461 citations)



2nd in CVPR 2017 NTIRE Image super-solution challenge



1st in ICDAR 2015 Blurring Text Classification challenge

SENSETIME
商汤科技

Diabetic retinopathy Diagnosis

Diabetic Retinopathy

- **Diabetic Retinopathy** is the leading cause of blindness in the working age population in developed countries. In 2015, it is estimated to affect over 93million people.

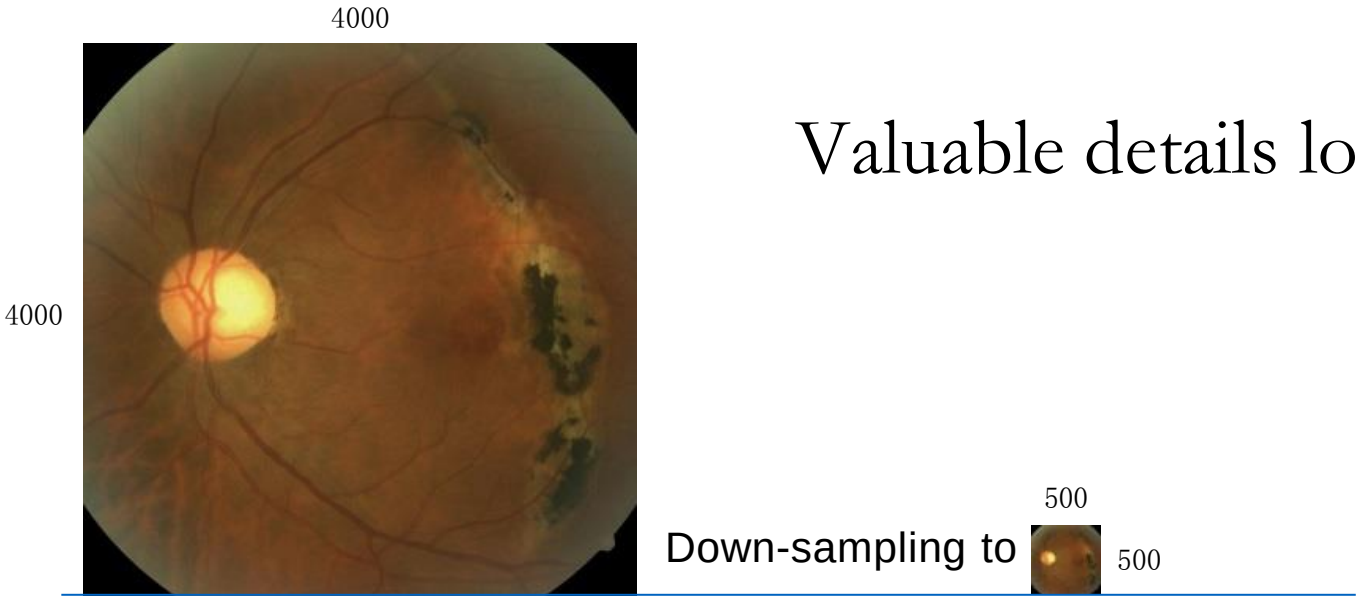
0:Normal 1:Mild 2:Moderate
3:Severe 4:Proliferative

- Google has published a paper on JAMA in 2017. They asked 54 ophthalmologists to label a total of 130 thousand images, and train a neural network with this dataset, which achieved a sensitivity of 90.3% and specificity of 98%.
- We proposed Zoom-in-Net achieved state-of-the-art performance on DR classification. Trained with only image-level labels, it can also accurate localize lesion regions with a recall score of 80% with 4 proposals per image.

Diabetic Retinopathy

Observation

- The size of a typical retinal image of fundus is about 4000*4000 ,while the size of a typical input image of a modern CNN is smaller than 500*500.
- The input size is limited by the GPU memory as well as the network complexity.

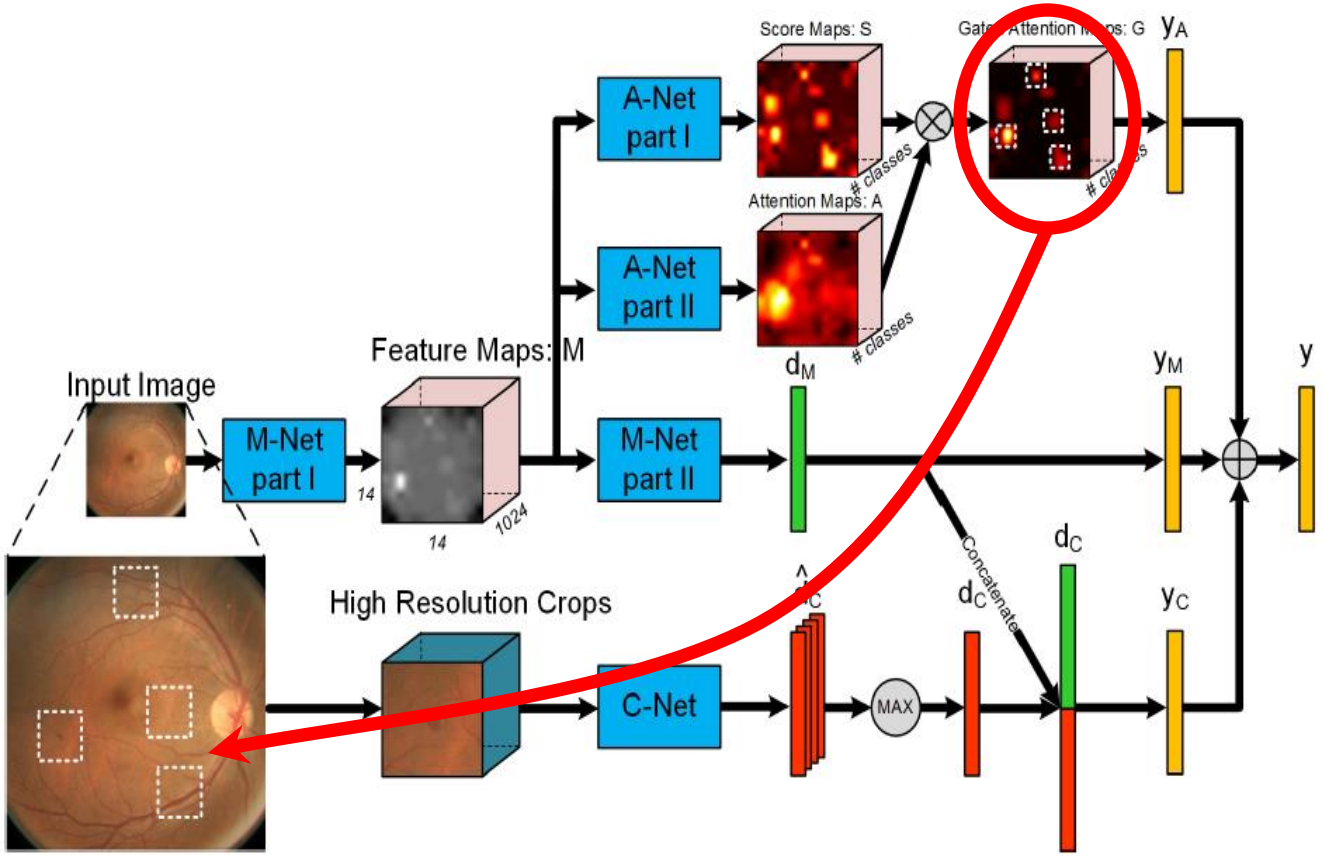


To get a sense of what the size difference is

Diabetic Retinopathy

Our Solution is Zoom-in-Net, which mimics the process of clinician’s behavior:

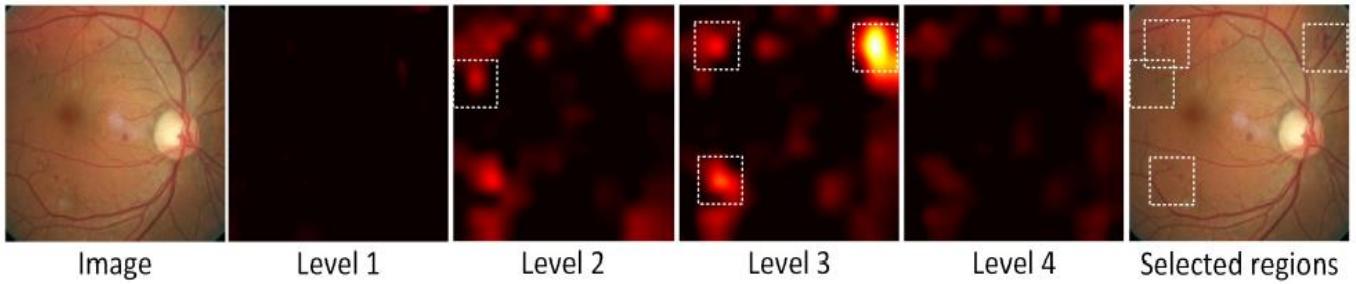
1. Scan the whole image to identify the suspicious regions
2. Automatically zoom in these regions to verify



Diabetic Retinopathy

Zoom in suspicious regions for details:

Sampling highest response regions on the gated attention maps in a greedy manner.



From left to right: image, gated attention maps of level 1-4 and the selected regions of the image.

Performance on Kaggle challenge and Messidor dataset

Algorithms	val set	test set
Min-pooling [45]	0.86	0.849
o_O [46]	0.854	0.844
Reformed Gamblers [47]	0.851	0.839
M-Net	0.832	0.825
M-Net+A-Net	0.837	0.832
Zoom-in-Net	0.857	0.849
Ensembles	**0.865**	**0.854**

Table 2.1: *Comparison to top-3 entries on Kaggle's challenge.*

Method	AUC	Acc.
Lesion-based [50]	0.760	-
Fisher Vector [50]	0.863	-
VNXK/LGI [48]	0.887	0.893
CKML Net/LGI [48]	0.891	0.897
Comprehensive CAD [51]	0.91	-
Expert A [51]	0.94	-
Expert B [51]	0.92	-
Zoom-in-Net	**0.957**	**0.911**

Table 2.2: *AUC for referable/nonreferable*

Method	AUC	Acc.
Splat feature/kNN [52]	0.870	-
VNXK/LGI [48]	0.870	0.871
CKML Net/LGI [48]	0.862	0.858
Comprehensive CAD [51]	0.876	-
Expert A [51]	0.922	-
Expert B [51]	0.865	-
Zoom-in-Net	**0.921**	**0.905**

Table 2.3: *AUC for normal/abnormal*

Diabetic Retinopathy

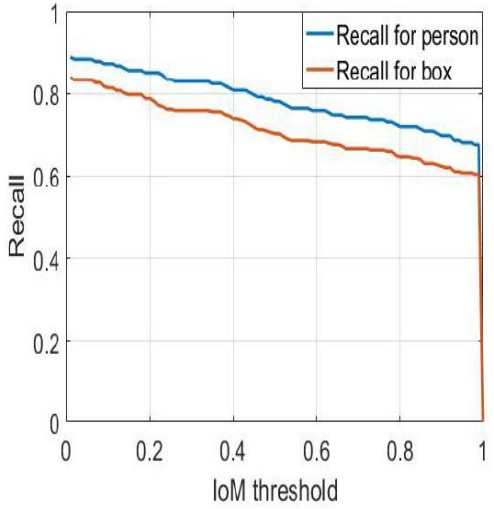
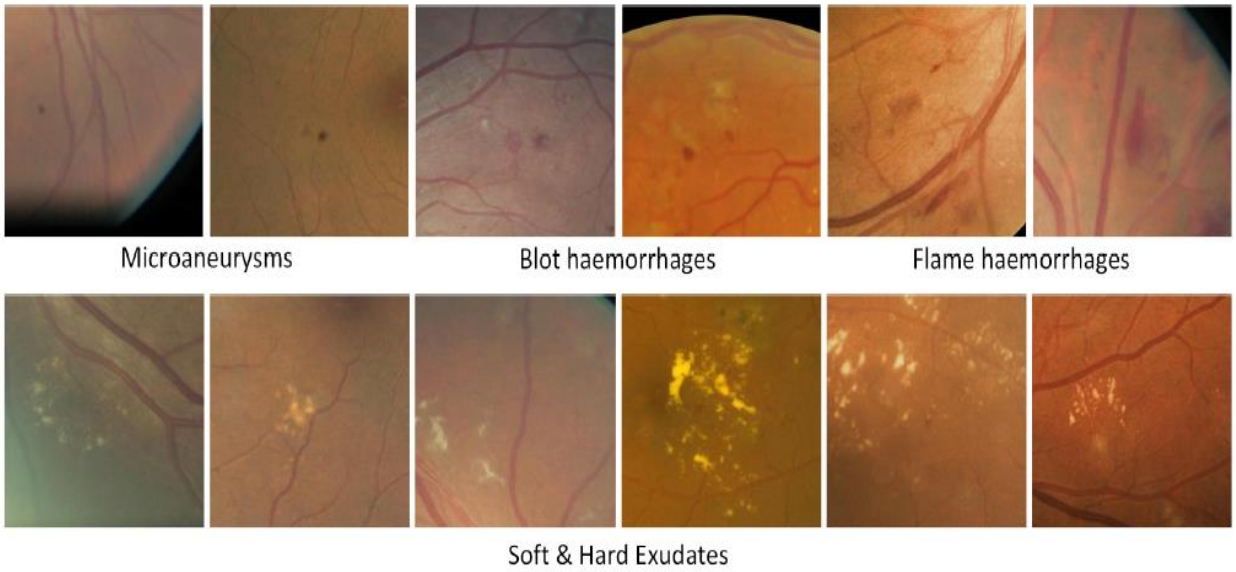


Figure 2.4: *AUC for normal/abnormal*



- Zoom-in-Net v.s. general physician v.s. optometrists[1]
  0.865 v.s. 0.838 v.s. 0.719
- 83% Recall for Zoom-in-Net with four proposals per image (left)
- Automatically discover meaningful lesion types (right)

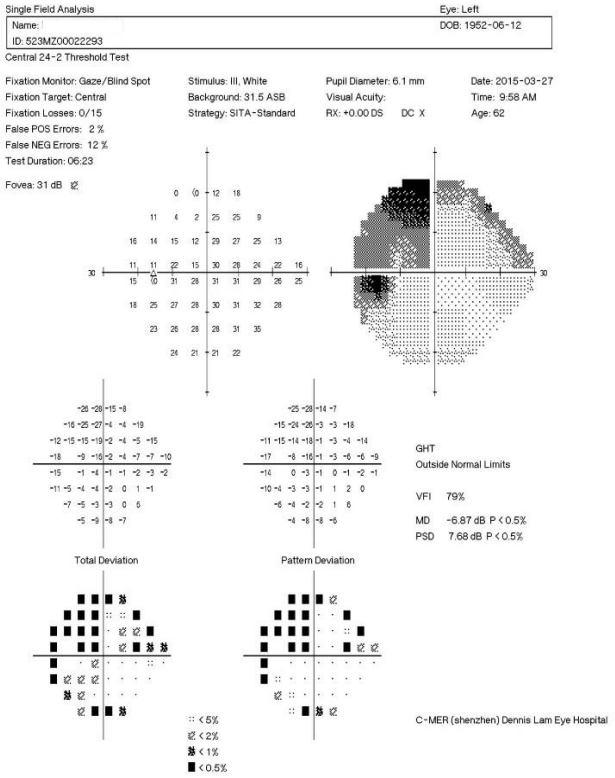
[1] Lalit Verma et al, "Screening for diabetic retinopathy by non-ophthalmologists: An effective public health tool", Acta ophthalmologica Scandinavica, 2013

SENSETIME
商汤科技

Glaucoma Diagnosis

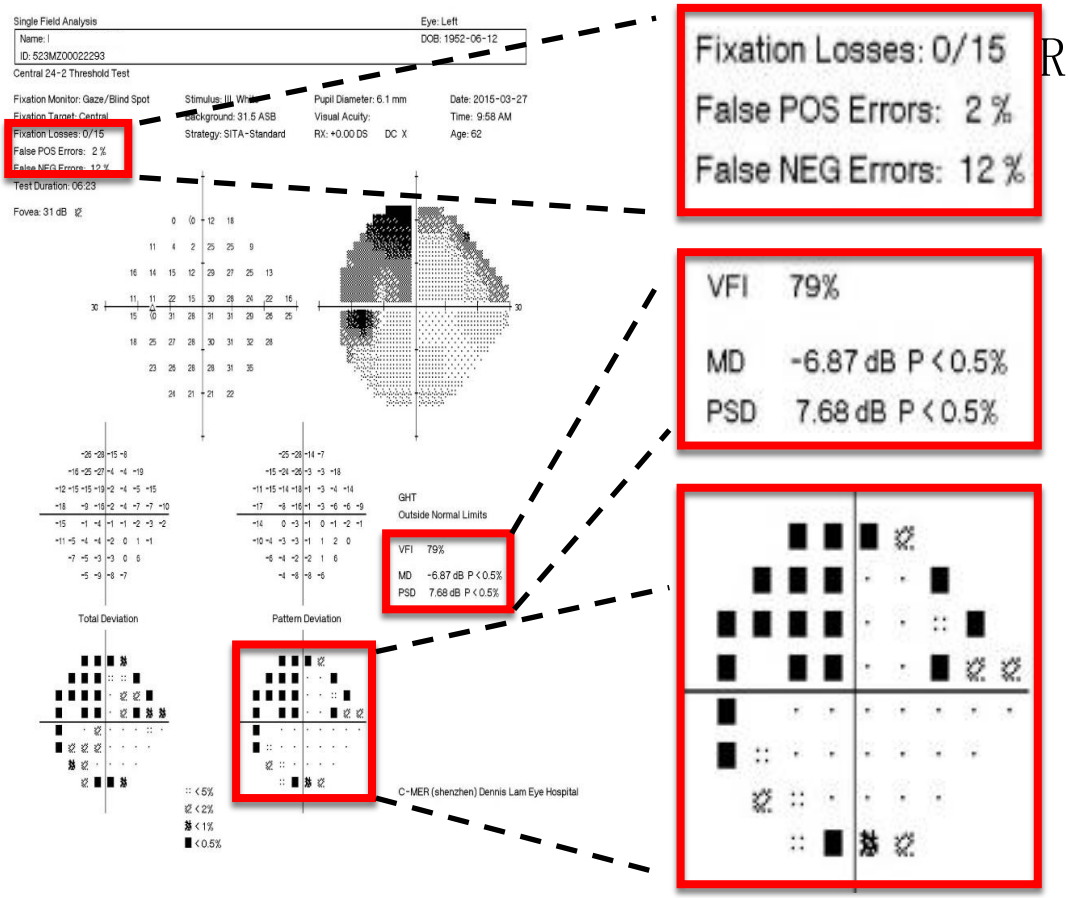
Glaucoma Diagnosis

- **Glaucoma diagnosis**
- Sensetime collaborated with Shenzhen Institute of Advanced Technology and ten ophthalmic hospitals.
- Over **7k** visual fields, **10k** OCT images and **10k** fundus images, the **largest** dataset for comprehensive diagnosis of glaucoma.
- Based on the visual reports, we designed a deep-learning-based algorithm which outperforms human experts.



Glaucoma Diagnosis

- Glaucoma diagnosis
- Our system can automatically verify the reliability of a visual field report by its fixation losses, False positive and False negative errors. And then it extracts the Pattern Deviation (PD) plot and process it with a convolution neural network.



Reliability check

Fixation Loss < 2/15

FP < 15%

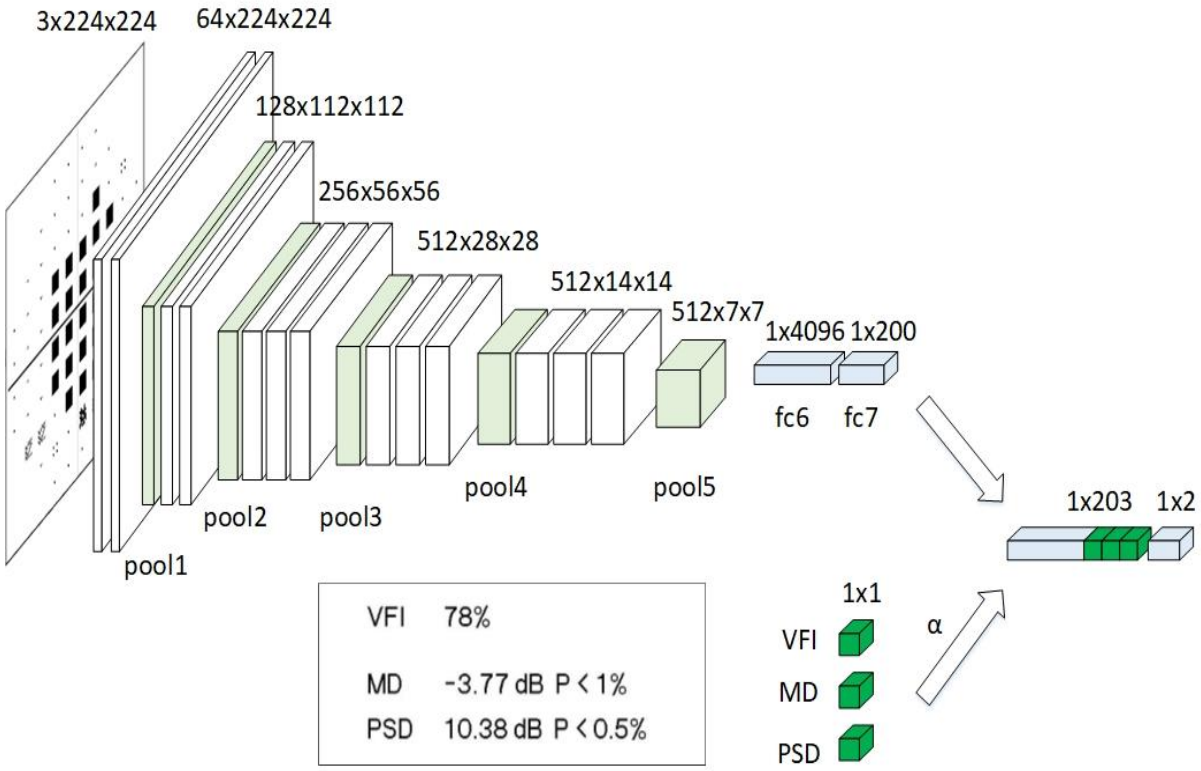
FN < 15%

Input of CNN

Input of CNN

Glaucoma Diagnosis

- Network structure
- The backbone of the network is Based on VGG16. We modify the output dimension of the penultimate layer fc7 from 4096 to 200 so that it is more comparable to the dimension of the three parameters VFI/MD/PSD.
- To fuse the potential information from VFI/MD/PSD, the 200-dimension output and the three parameters are concatenated to form a fused vector, which is used for final prediction.

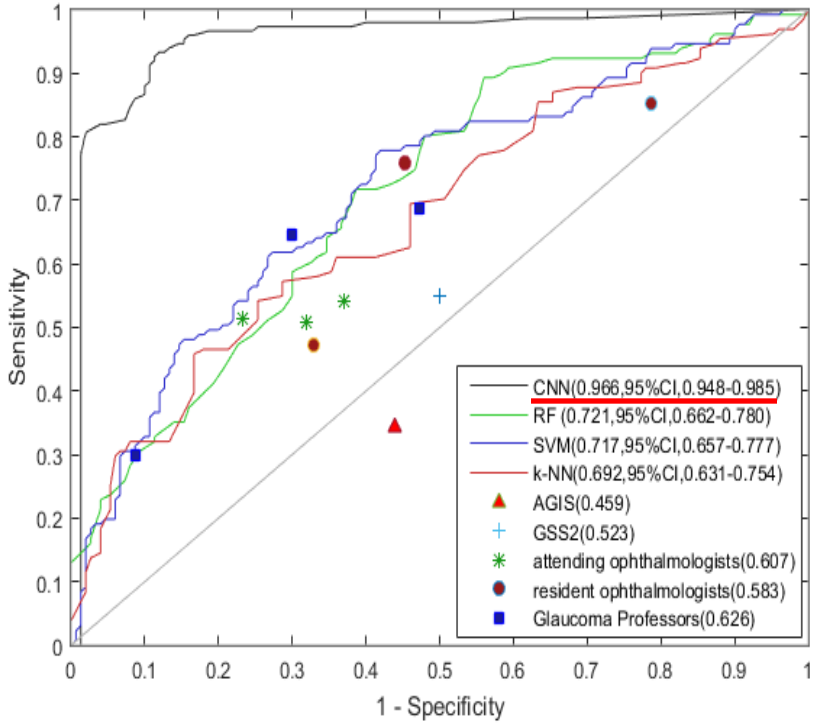


Glaucoma Diagnosis

- Glaucoma diagnosis
- Left figure shows baseline statistics
- Right figure shows the performance on a test dataset of 300 PD plots, compared to traditional approaches and nine ophthalmologists.

	Non-glaucoma Group	Glaucoma Group	*P Value
No. of images	1623	2389	-
age	47.2 (17.4)	49.2 (16.3)	0.0022
left/right	635/919	607/911	0.6211
VFI	0.917 (0.126)	0.847 (0.162)	0.0001
MD	-5.0 (23.5)	-9.0 (44.8)	0.0039
PSD	3.6 (3.3)	6.7 (22.2)	0.0001

Mean and standard deviation are provided. *Comparison between normal group and glaucoma group (unpaired t test for numerical data and chi-square test for categorical data)

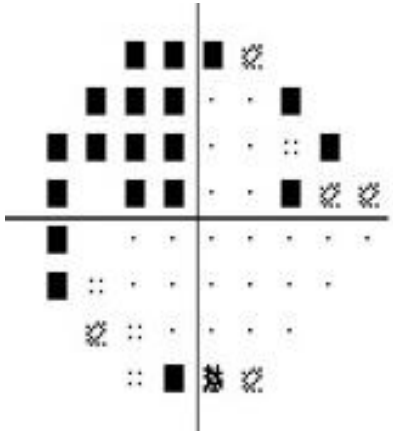


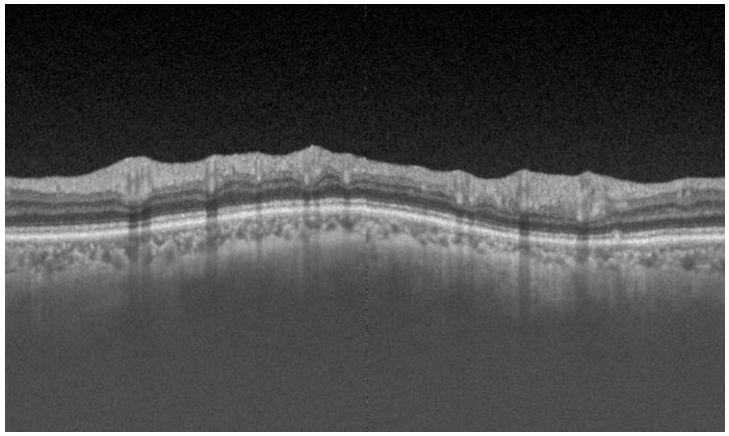
Glaucoma Diagnosis

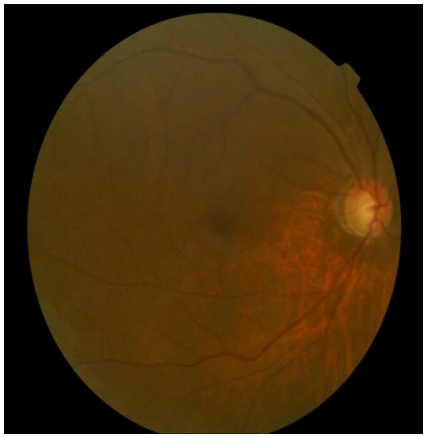
- Now we are developing **a cloud-based platform** for clinical trials.

We hope this A.I. system can really help improve the diagnosis accuracy

- We are also working on integrating OCT images and fundus images in the system.
- Different modalities, better performance







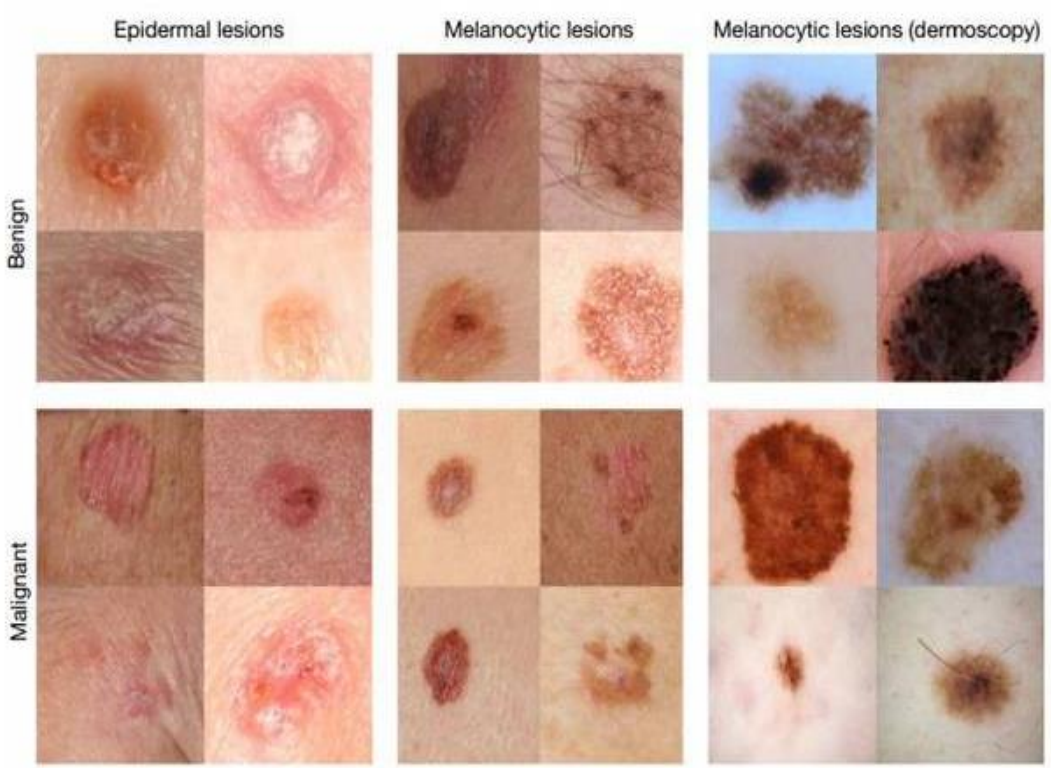
Visual Field + Fundus Image + OCT

SENSETIME
商汤科技

Skin Lesion Analysis

Skin Lesion Analysis

- **Skin lesion diagnosis**
- With big data and deep learning, A.I. can perform as good as human experts[1].



- Currently, our system outperforms the state-of-the-art methods on ISIC Skin challenge datasets on both the classification task and segmentation task.

1. Dermatologist-level classification of skin cancer with deep neural networks, Nature 2017

Skin Lesion Analysis

- **Skin lesion** classification and segmentation are two highly correlated tasks. However, their relationship is not fully utilized in previous methods.
- We designed a multi-task deep convolution neural network architecture to solve both tasks simultaneously.
- A feature passing module is proposed for message transmission between the two networks.

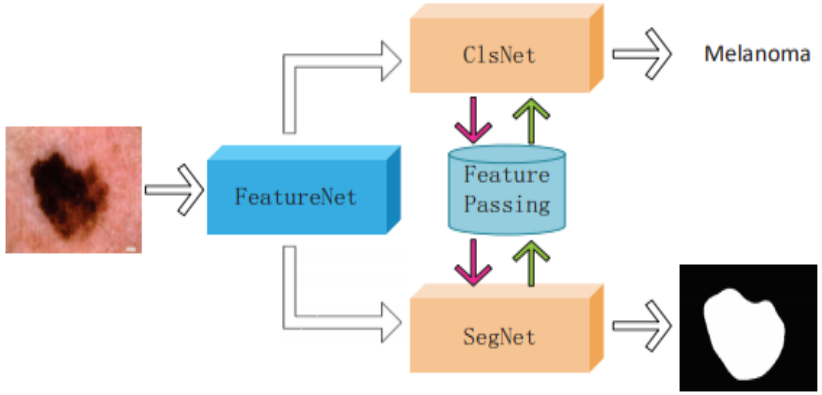


Fig. 1. Workflow of the proposed algorithm

- Currently, our system achieves the accuracy of 81% and a mIoU of 0.79 on the ISIC challenge (both best).

1. Dermatologist-level classification of skin cancer with deep neural networks, Nature 2017

Skin Lesion Analysis

Table 1. Performances compared with our base model and multi-task network without feature passing module.

Method	AC(cls)	mAP(cls)	JA(seg)	AC(seg)	DI(seg)
Base	0.772	0.699	0.779	0.940	0.862
Multi-task	0.779	0.712	0.780	0.940	0.863
Ours	0.801	0.747	0.787	0.944	0.868

Table 2. Classification performance compared with other state-of-art classification models.

Method	AC(cls)	mAP(cls)
AlexNet	0.775	0.706
VGG16	0.789	0.727
ResNet101	0.772	0.699
Ours	0.801	0.747

Table 3. Segmentation performance compared with other state-of-art segmentation models.

Method	JA(seg)	AC(seg)	DI(seg)
MtSinai	0.765	0.934	0.849
U-net	0.741	0.926	0.822
Deeplab-ResNet101	0.779	0.940	0.862
Ours	0.787	0.944	0.868

SENSETIME
商 汤 科 技

Conclusions

Deep learning for human vision in many ways

- Better view of photos (enhancement, super-resolution, …)
- Diagnosis of eye related disease (Glaucoma, Diabetic Retinopathy,…)
- Scan huge image datasets (medical, surveillance, …)

SENSETIME 商汤科技

SENSETIME 商汤科技

THANK YOU!

Follow Sensetime on WeChat