低度訓練資源的深度學習方法及產業應用範例

林嘉文 (Chia-Wen Lin)
清華大學電機系教授 & AI研究中心副主任
cwlin@ee.nthu.edu.tw

6/9/2020

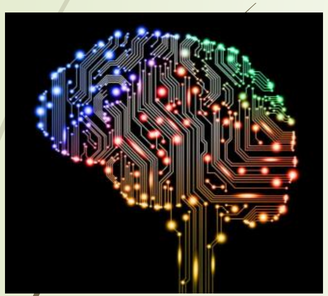
Outline

- **AI & Deep Learning**
- Low-Shot Learning
- Application Examples
  - Domain-Transferred Face Augmentation
  - IC-Fab Process Simulation
- Summary

ABCD for AI

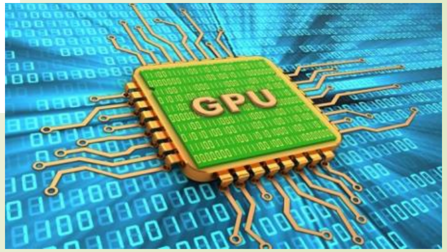
Steering Wheel (Business)

Fuel (Data)



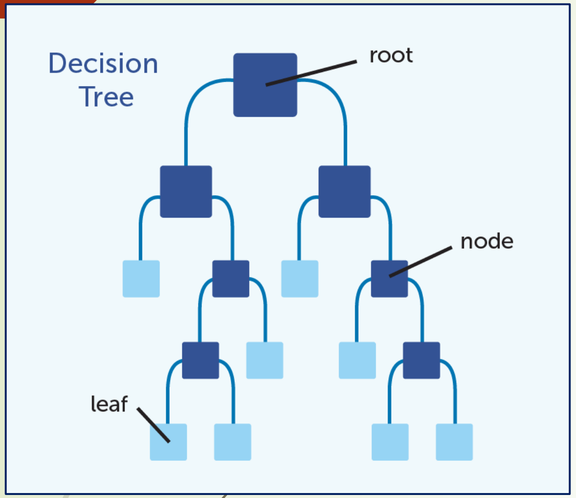
Engine (Algorithm)

Wheels (Computing)

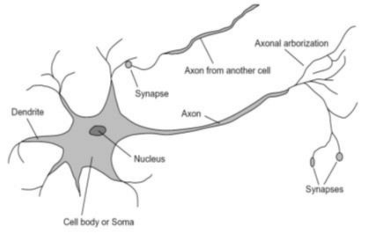


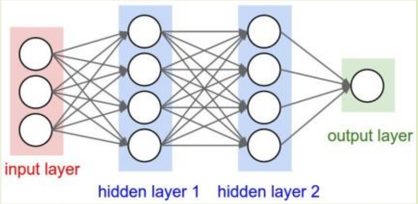
Schools of Machine Learning **Algorithms**

Decision Tree



Neural Net





Symbolists

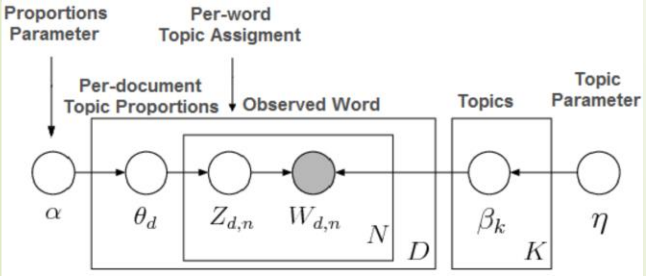
Connectionists

Bayesians

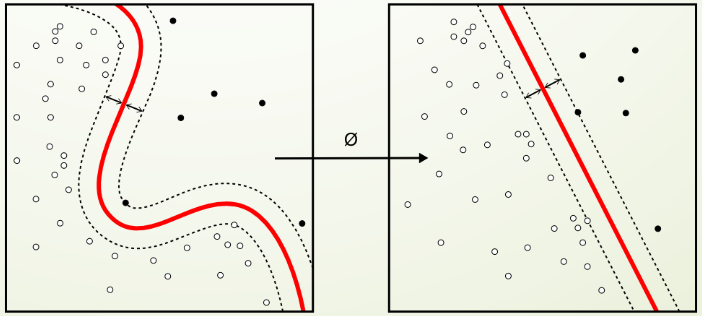
Analogizers

Evolutionists

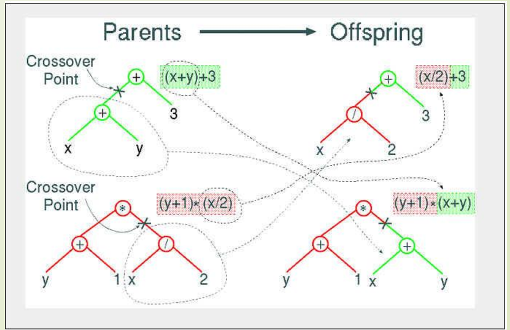
Probabilistic Model



SVM

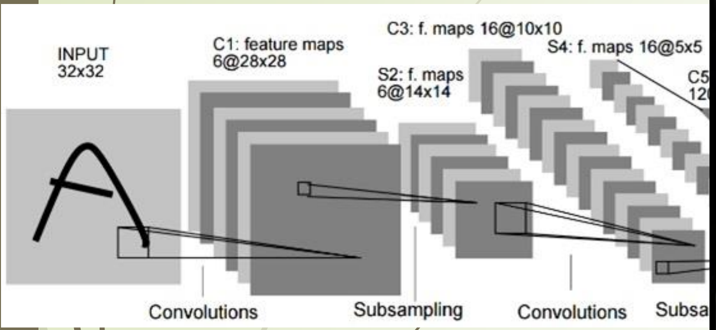


Evolutionary Computing

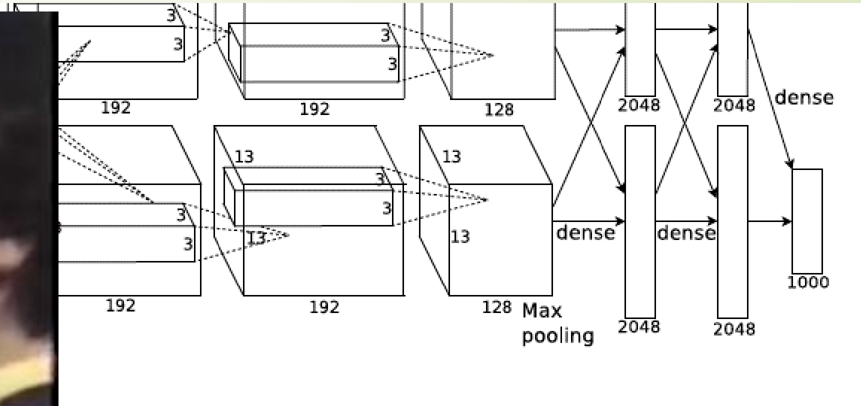


Connectionists: CNNs

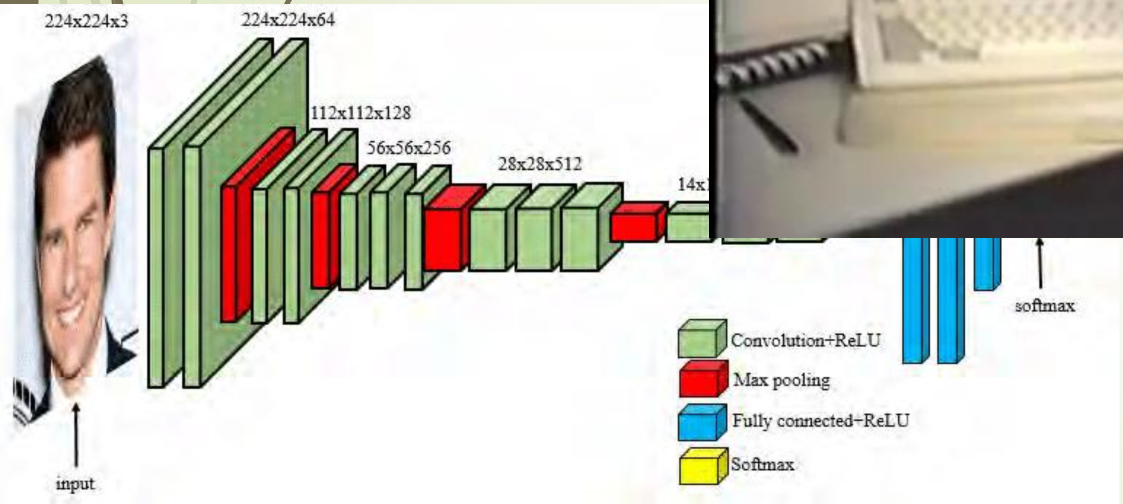
LeNet 1998



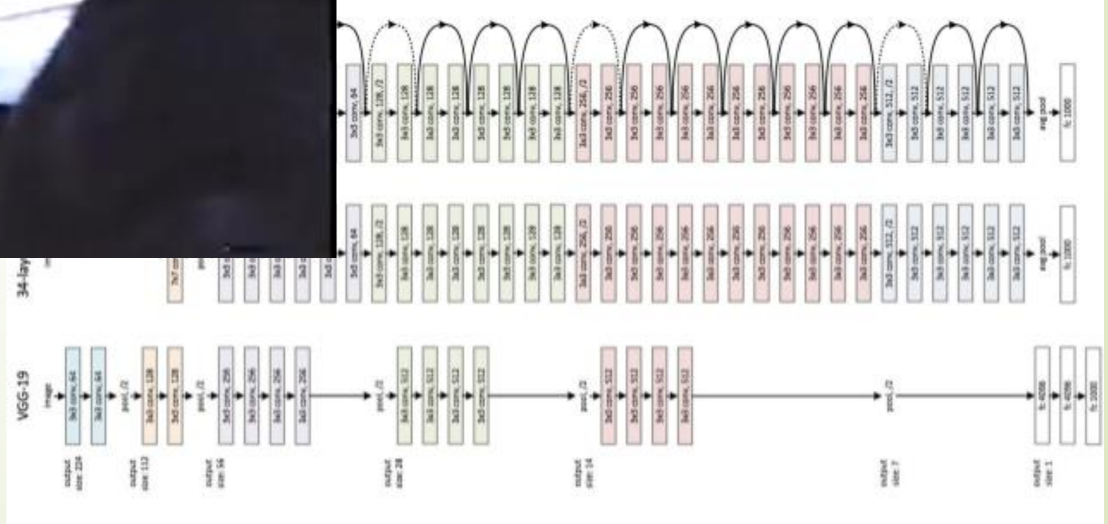
AlexNet (Supervision) 2012



VGG 2014



ResNet 2015

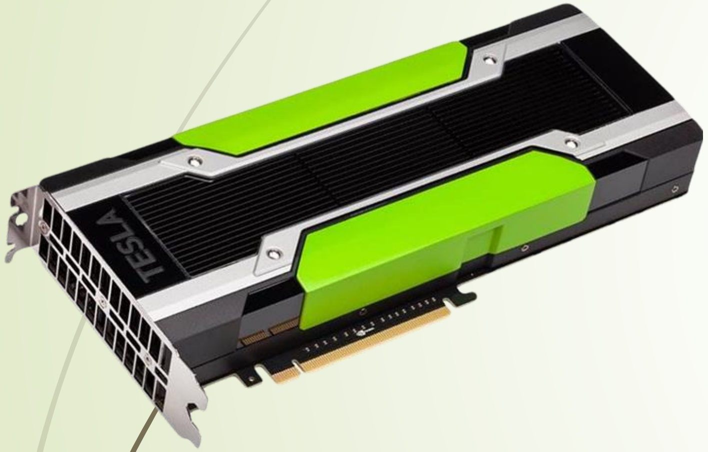


Connectionists: CNNs

		place	top 5 error
2011	Compressed Fisher kernel + SVM	1st	25.8%
2012	SIFT + GIST + LBP + PA classifier	2nd	26.1%
2012	Supervision	1st	16.4%
2013	Clarifai	1st	11.5%
2014	VGG	2nd	7.3%
2014	GoogLeNet / Inception	1st	6.6%
2014	*Andrej Karpathy* 	n/a	5.1%
2015	Batch Normalization Inception	n/a	4.8%
2015	Inception v3	2nd	3.6%
2015	ResNet	1st	3.6%
2016	Inception-ResNet	n/a	3.1%

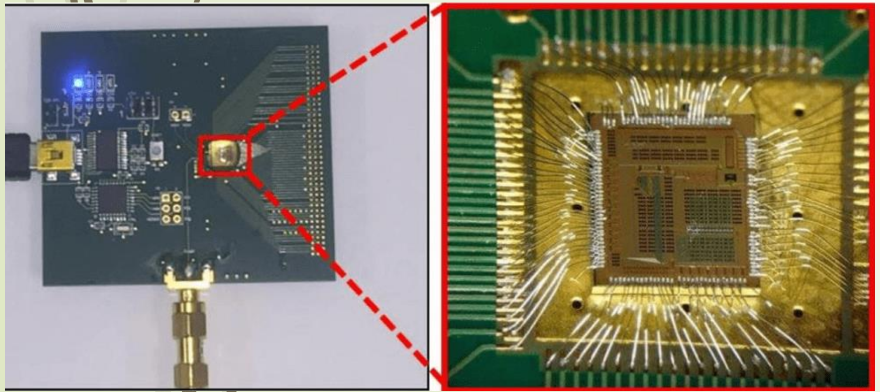
Computing for AI

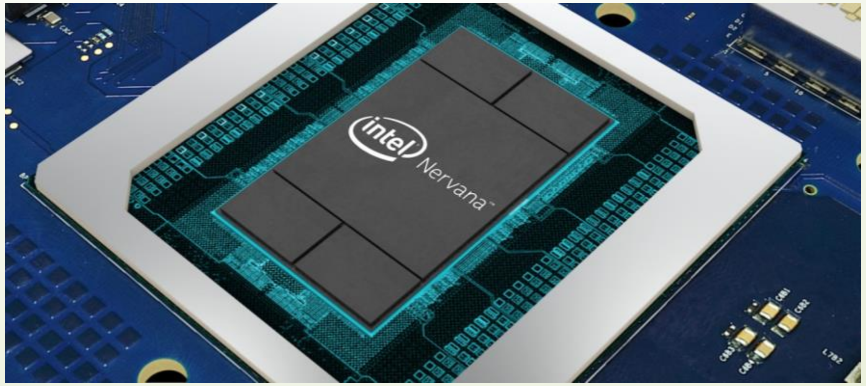
Graphics Processing Unit (GPU)

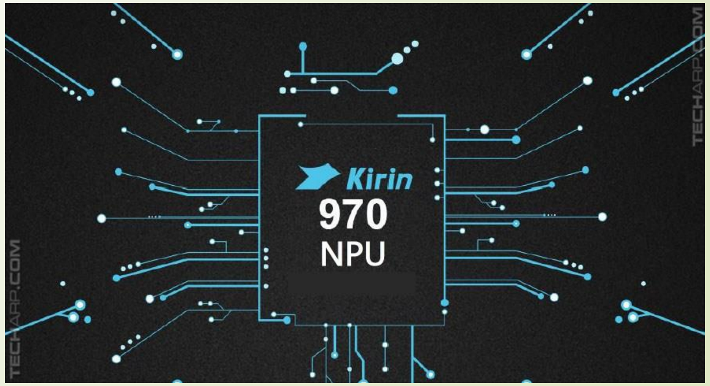


Tensor Processing Unit (TPU)

Neural Processing Unit (NPU)







Data for AI


BIG
DATA
INFORMATION
LARGE
STORAGE
SETS
SEARCH
PETABYTES
ANALYSIS
RESEARCH
MANAGEMENT
SOFTWARE
TECHNOLOGIES
VOLUME
PROCESS
DEVELOPMENT
TECHNOLOGY
MILLION
SCIENCE
INTERNET
FUTURE
BILLION
NEW

Data for AI

Era of Data Exploration

- 90% ----- Currently, 90% data were generated in recent two years
- 1.7 MB ----- In 2020, each person will generate 1.7 MB per second on average
- 50 B ----- In 2020, there will be ≥ 50 B Internet devices

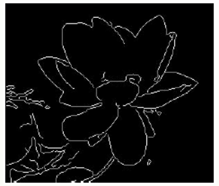
Status of Data Usage

- 0.5% ----- The current percentage of data analyzed/used
- 10% vs. 65M ----- We can obtain 65M revenue for every increased 10% data usability

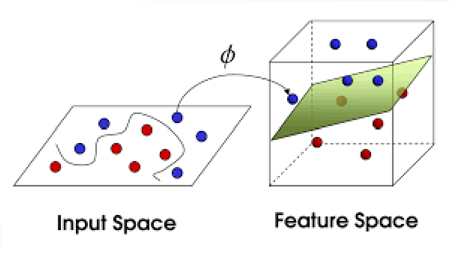
Conventional Approach to Object Recognition

- Training phase

image collection → feature extraction → classifier training → trained classifier



histogram of oriented gradients (HoG)



support vector machines (SVMs)

dogs? ✖

flowers? ○

trains? ✖

persons? ✖

Conventional Approach to Object Recognition

- Training phase

image collection → feature extraction → classifier training → trained classifier

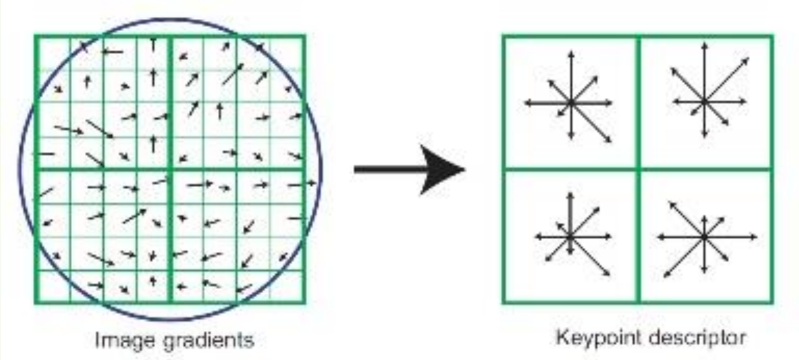
- Testing phase

test image → feature extraction → trained classifier → prediction

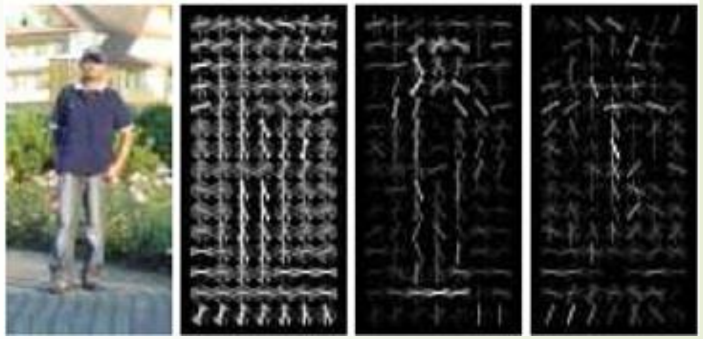
→ people

Features are The Keys

- Off-the-shelf visual features

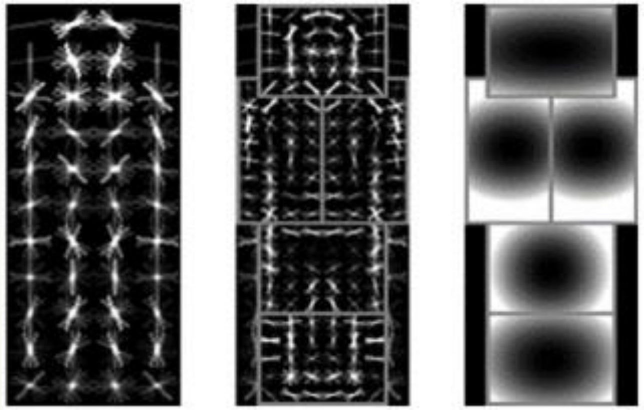


SIFT [Lowe, IJCV'04]
Citations: 43465



HoG [Dalal & Triggs, CVPR'05]
Citations: 20174

Constellation model [Fergus et al., CVPR'03]
Citations: 2551



DPM [Felzenszwalb et al., PAMI'10]
Citations: 5093

Conventional Approaches vs. Deep Learning

- Conventional approaches
  - Fixed/engineered features + trainable classifier

Hand-crafted Feature Extractor → "Simple" Trainable Classifier →

- Deep learning / End-to-end learning / Feature learning
  - Trainable features + trainable classifier

Trainable Feature Extractor → **Trainable** Classifier →

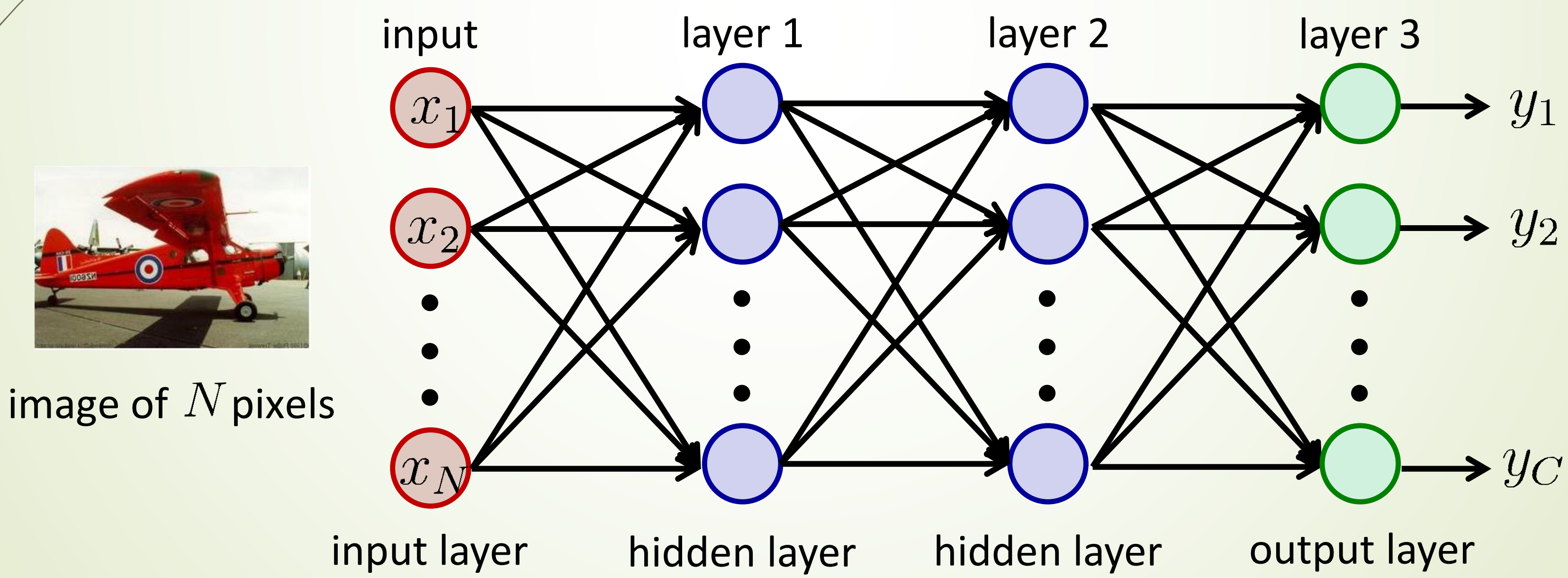
slide: Y LeCun & MA Ranzato

Deep Learning = Learning Hierarchical Rep.

- Each layer of hierarchy extracts features from output of previous layer
- All the way from pixels → classifier
- Layers have the (nearly) same structure

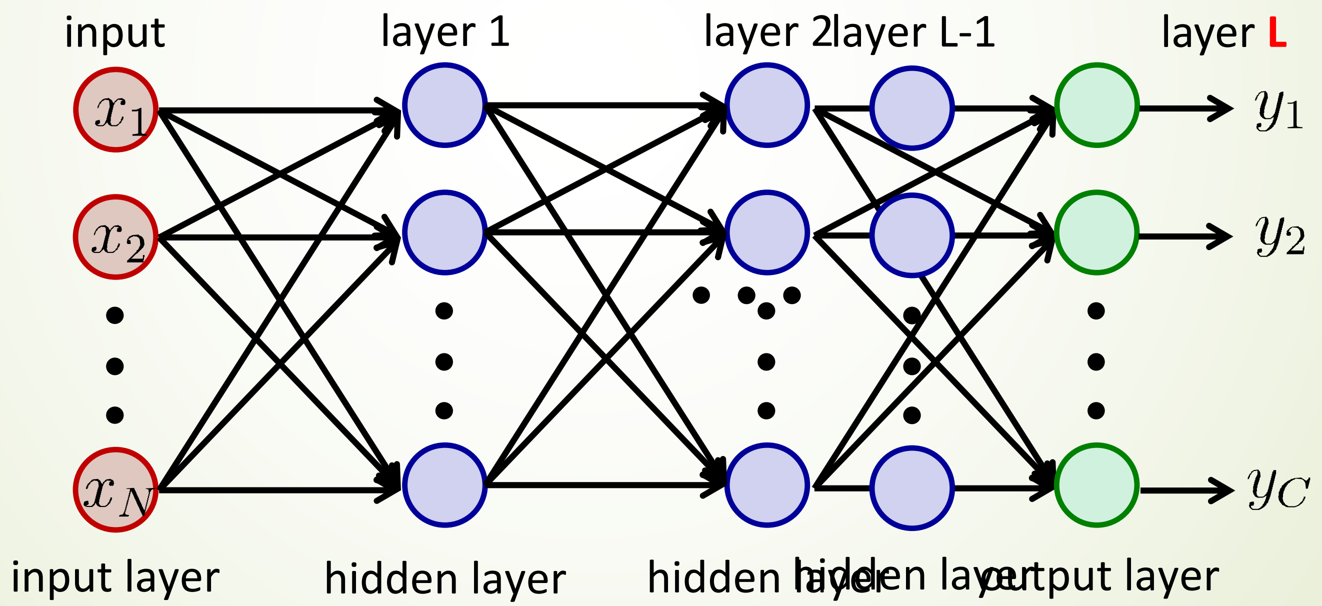
Neural Networks and Neurons

- Neural networks are presented as layers of interconnected neurons
  - Each layer of neurons takes messages from output of previous layer



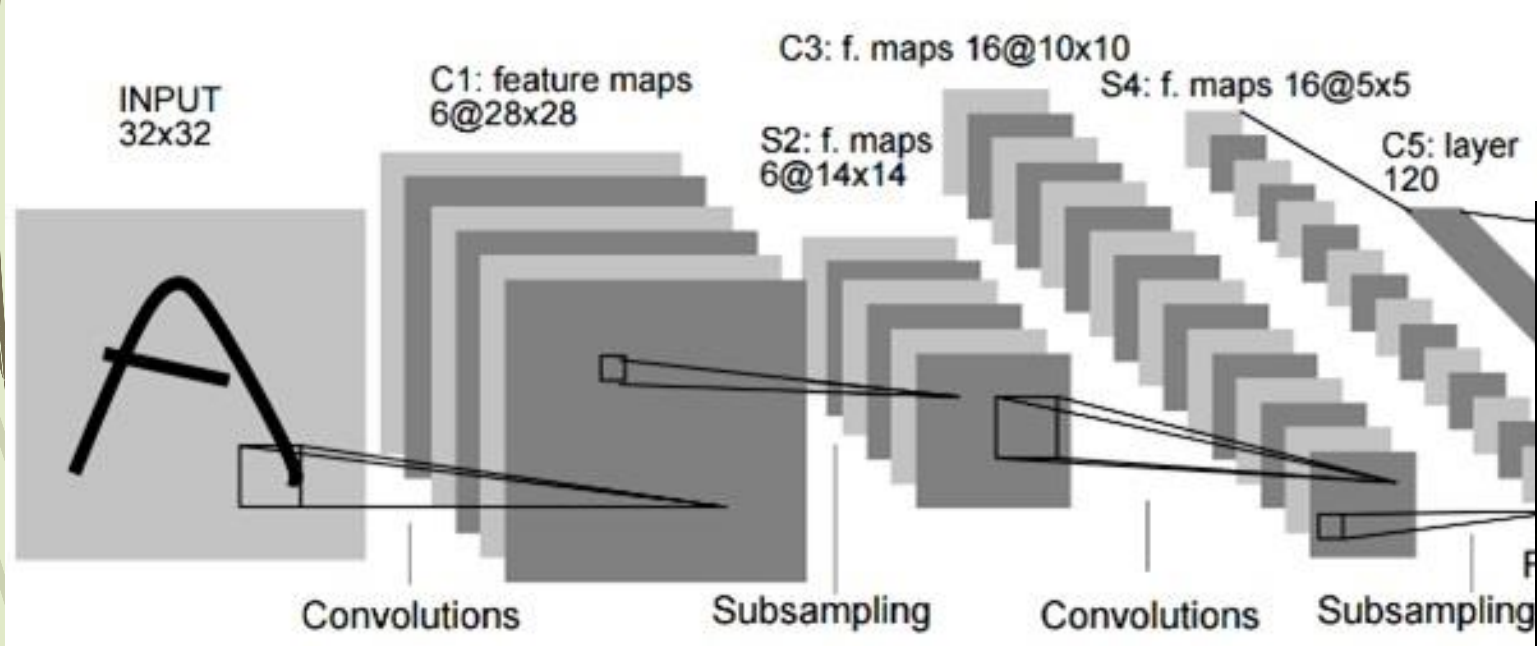
What are Deep Neural Networks (DNNs)

- DNN is neural networks with many hidden layers



Pioneer of CNNs: LeNet (1998)

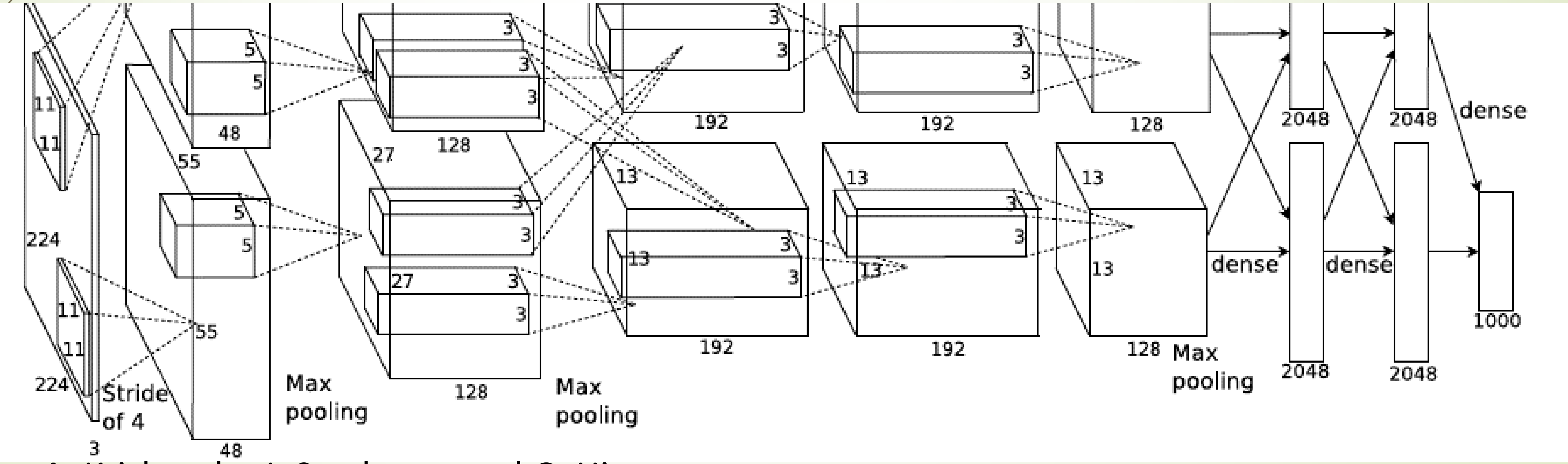
Yann LeCun



AlexNet (2012)

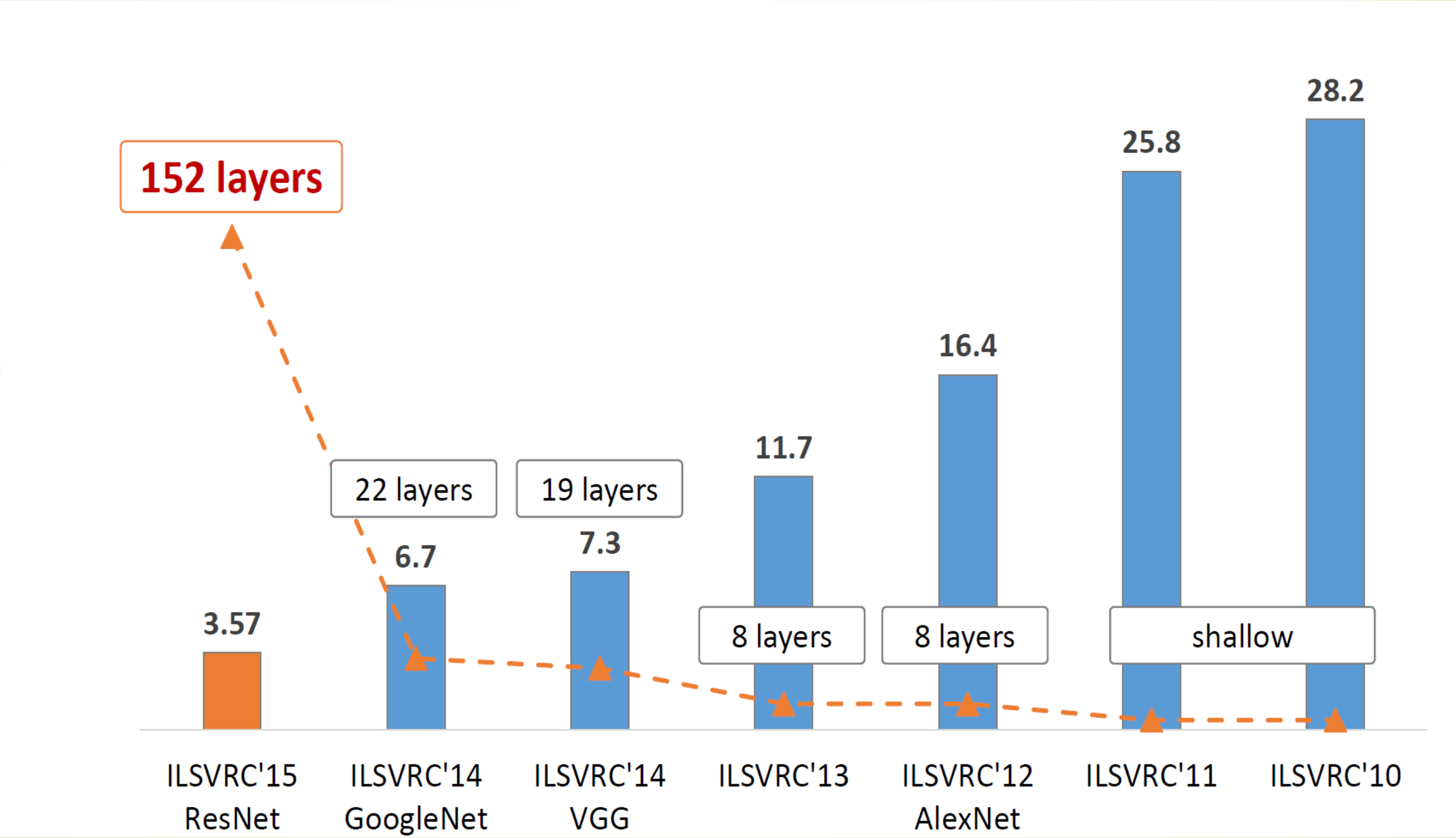
- Similar framework to LeCun'98 but:
  - Bigger model (7 hidden layers, 650,000 units, 60,000,000 params)
  - More data (10^6 vs. 10^3 images)
  - GPU Implementation (50x speedup over CPU)
    - Trained on two GPUs for a week

Geoffrey Hinton



A. Krizhevsky, I. Sutskever, and G. Hinton,
ImageNet Classification with Deep Convolutional Neural Networks, NIPS 2012

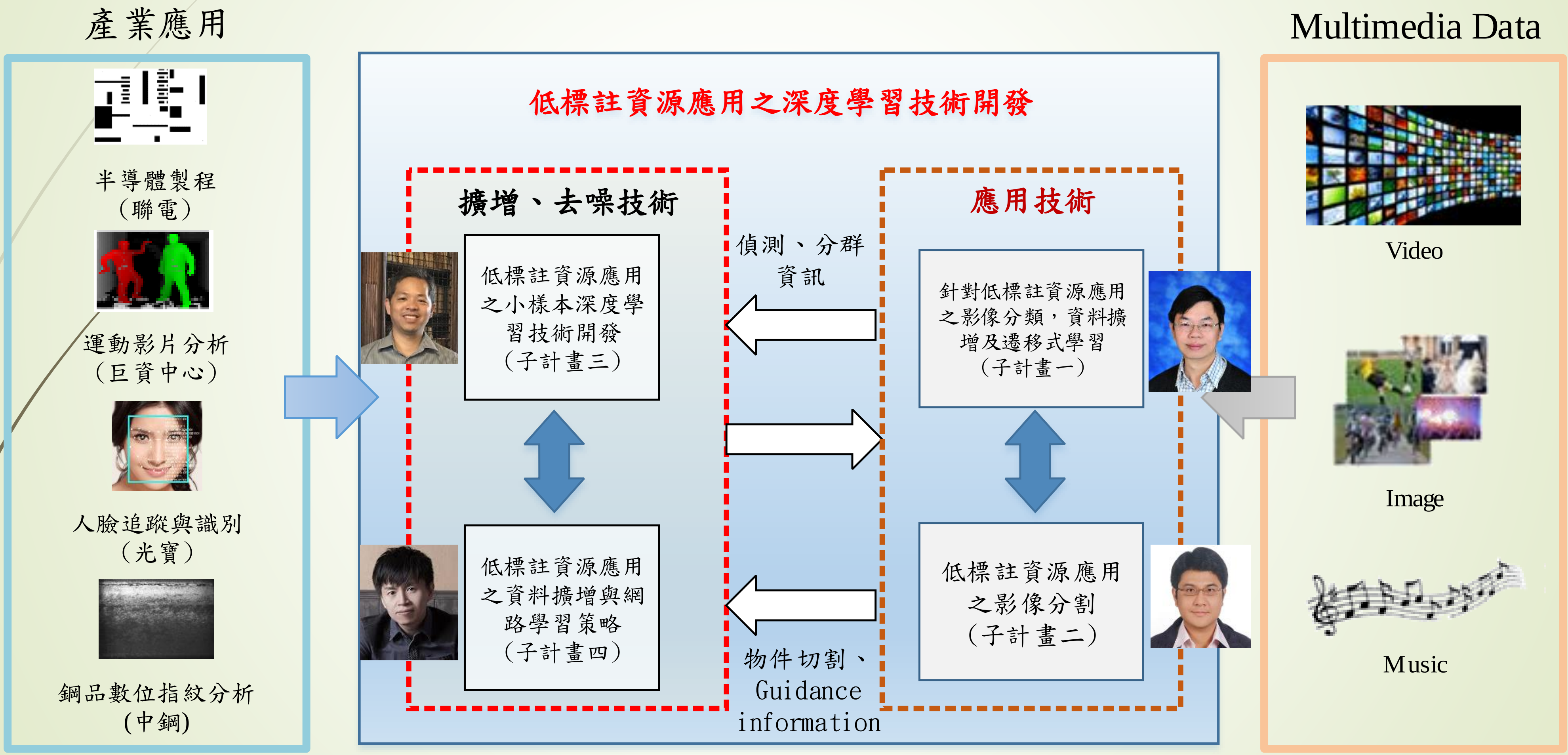
Evolution of # of Layers of CNNs



Outline

- AI & Deep Learning
- **Low-Shot Learning**
- Application Examples
  - Domain-Transferred Face Augmentation
  - IC-Fab Process Simulation
- Summary

科技部AI計畫：低標註資源應用之深度學習技術開發



Problems in AI & Deep Learning

Recent Significant Process on AI was mainly on DL

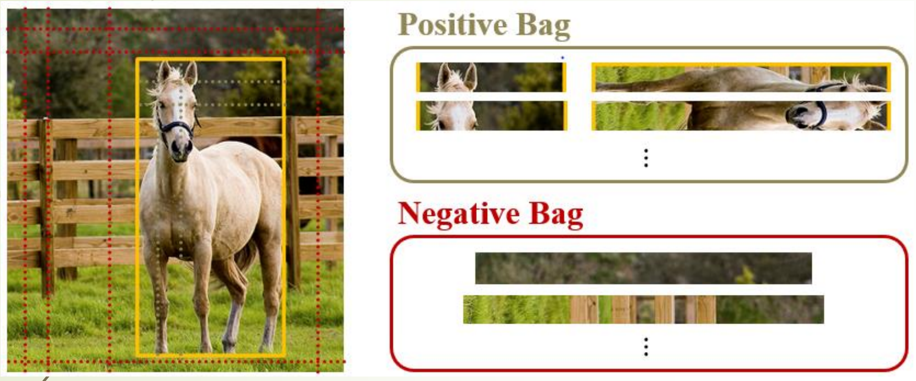
- in **supervised learning** applications
  - basically kind of curve fitting
  - requires **comprehensive & good-quality labelled** training data
- Many applications can only use **"weakly-supervised"**, **"semi-supervised"** & **"non-supervised"** learning
  - Low-shot learning

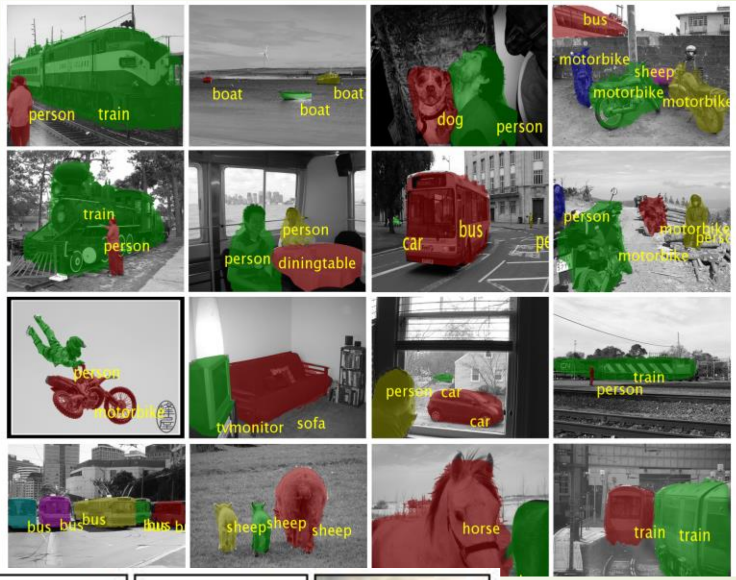
Low-Shot (few-, one-, zero-shot) Learning Problems:

- Low Labeling Resource (amount & information)
- Small Training Set
- Unbalanced (Biased) Training Data
- Label Noise (cloud-sourcing, different professional opinions)
- Domain Shift

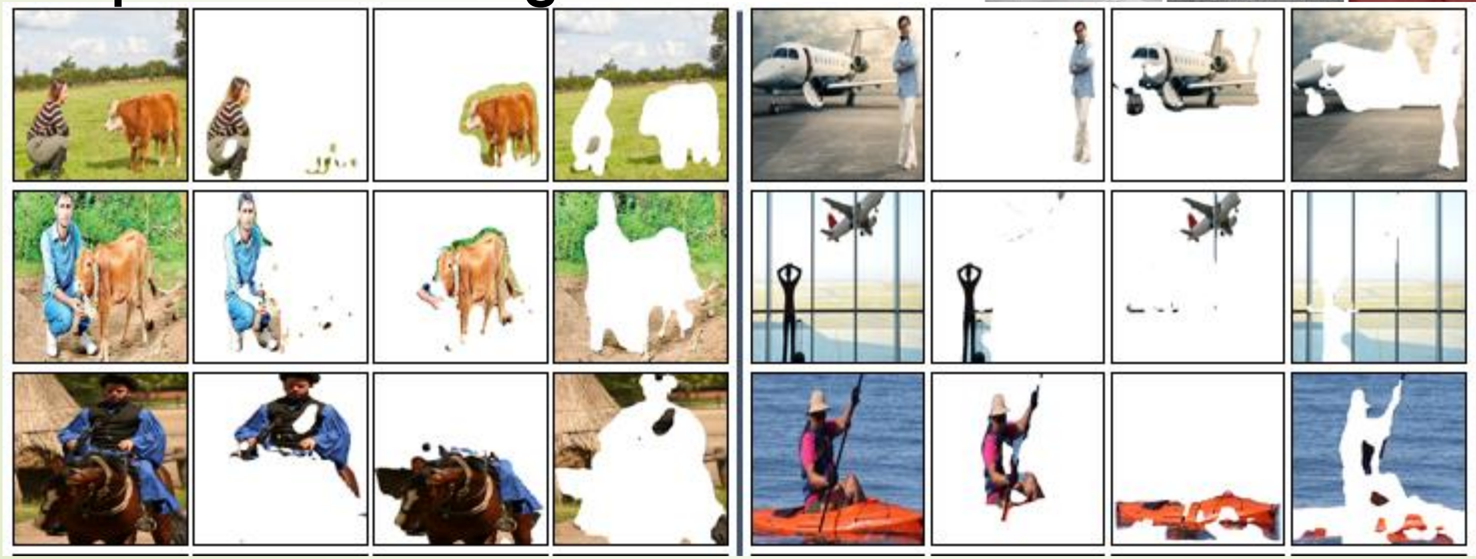
Low-Shot Learning for Semantic Segmentation

semi-supervised learning





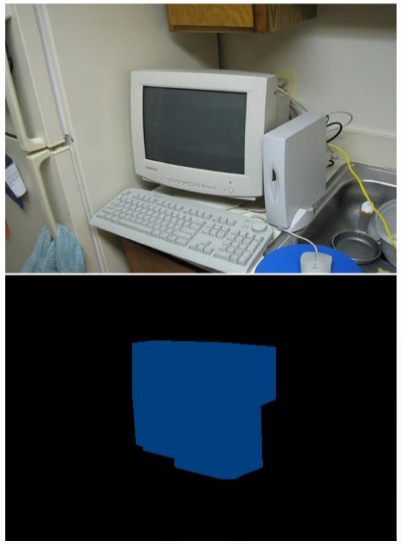
unsupervised learning



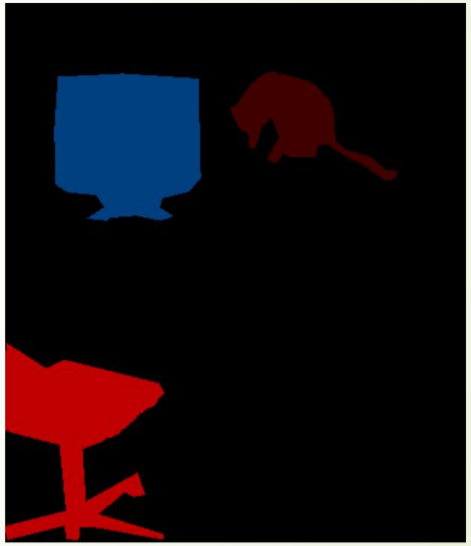
One/Few-Shot Learning

Learn a concept from one or only a few training example

Input



Novel Class



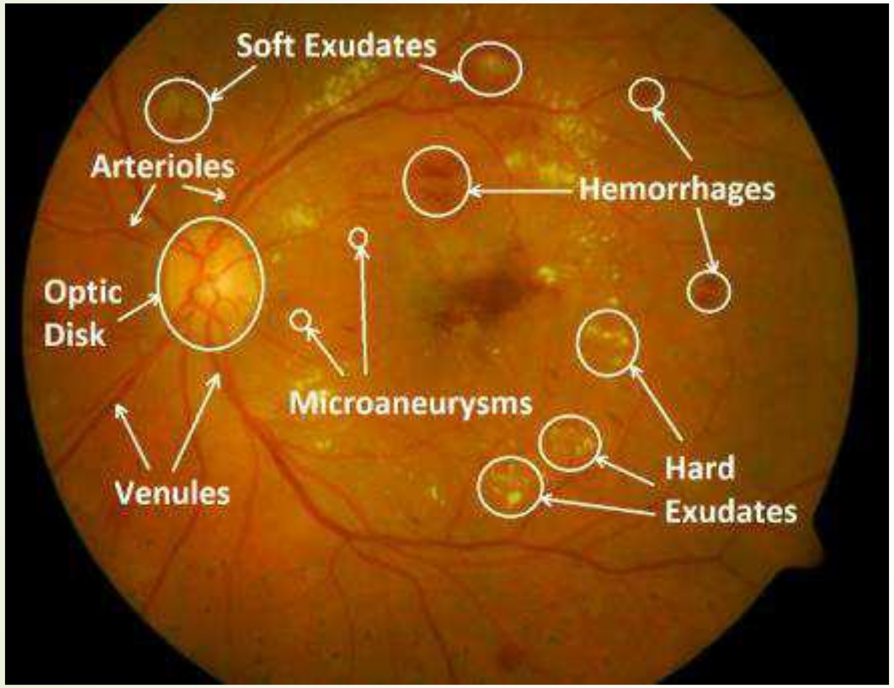
Result

One/Few-Shot Learning

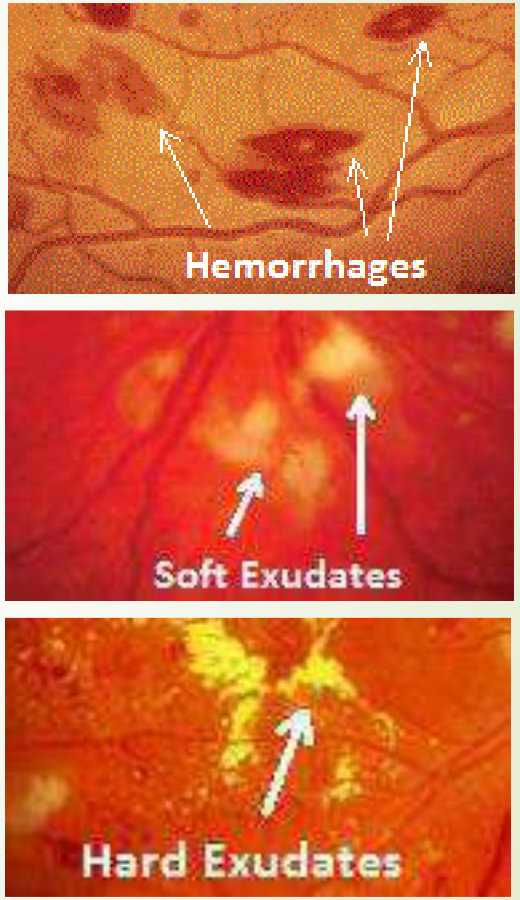
Person Re-identification

Low-resolution | shade | Pose variation | illumination | Blurred vision

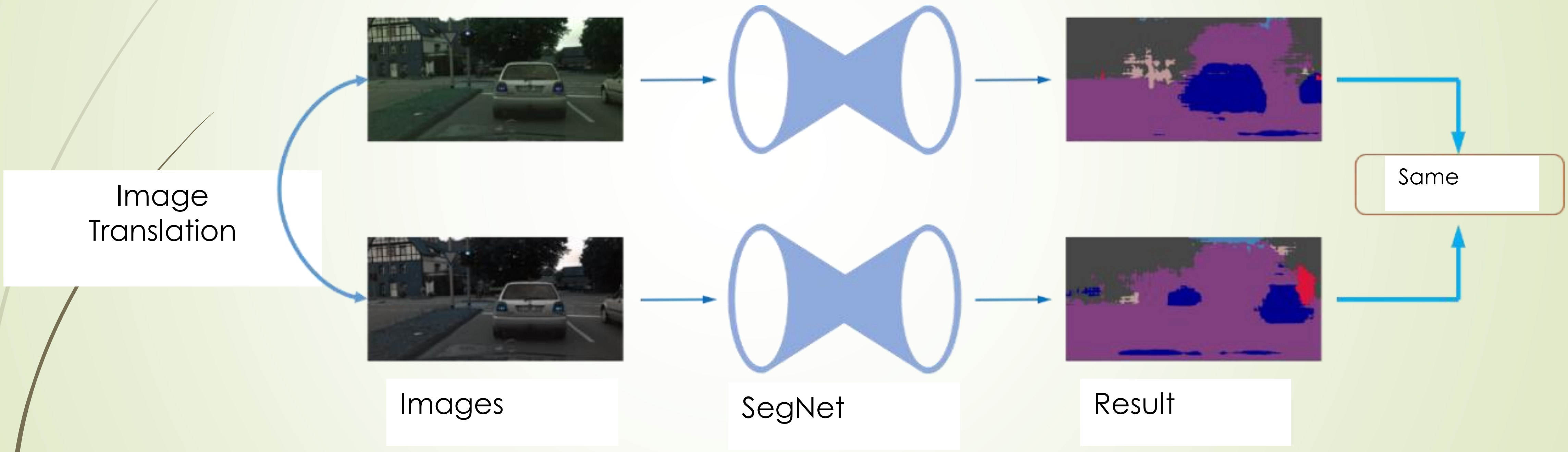
Retinal Image Analysis for Diabetes Eye Diseases Detection & Recognition



target



Transfer Learning

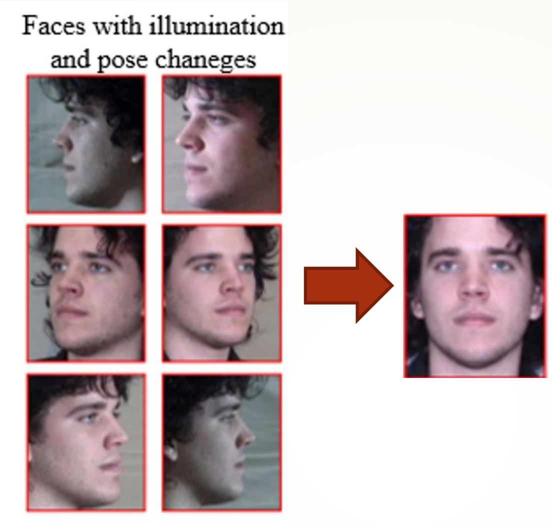


Outline

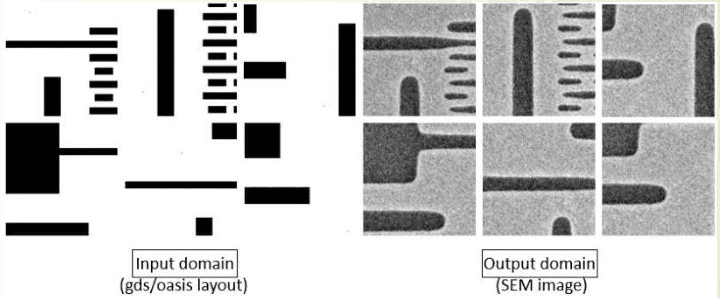
- AI & Deep Learning
- Low-Shot Learning
- **Application Examples**
  - **Domain-Transferred Face Augmentation**
  - **IC-Fab Process Simulation**
- Summary

Application Examples

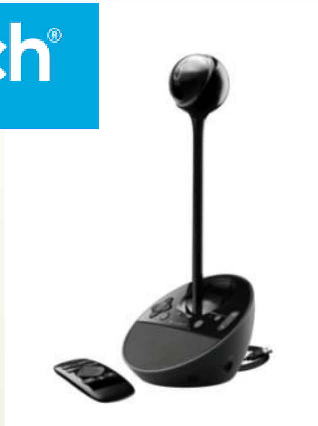
LITEON®
光寶集團
LITE-ON GROUP



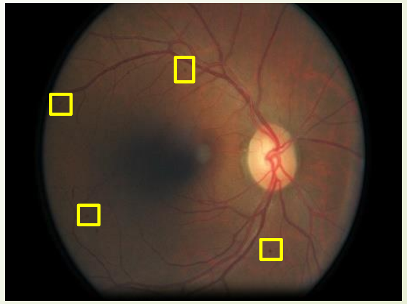
UMC
聯 華 電 子



logitech®



工業技術研究院
Industrial Technology
Research Institute



Industry Application
Face Augmentation for Low-Shot Recognition

Chia-Wen Lin (林嘉文), IEEE Fellow

Professor, Department of Electrical Engineering

National Tsing Hua University, Taiwan

cwlin@ee.nthu.edu.tw

Motivation: Difficulties in Face Recognition

- In surveillance applications, there often exist environmental interferences:
  - **Illumination variations**
  - **Pose/Viewpoint changes**
  - **Facial expressions**
  - Occlusions (eyeglasses, masks)

Pre-trained Face Recognition Model: LightCNN

Feature representation 1

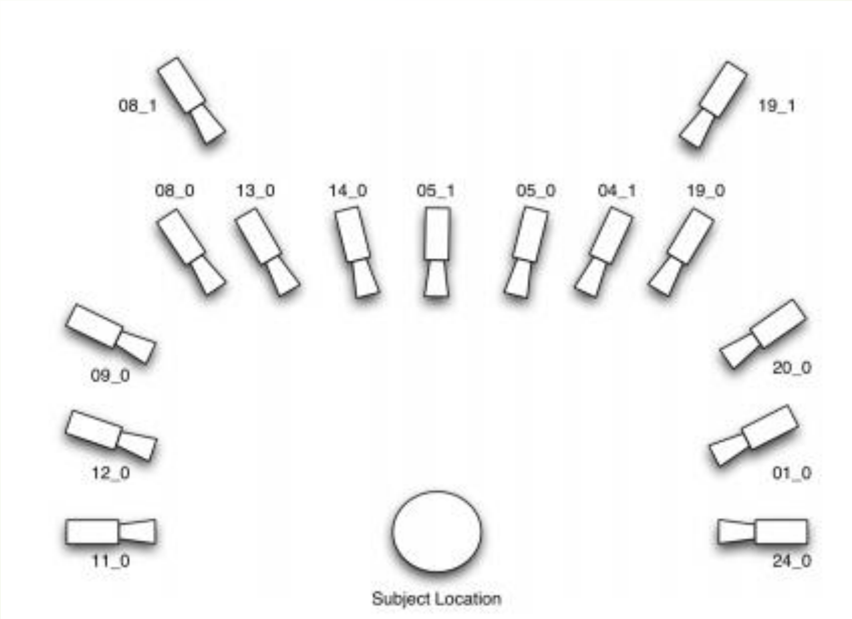
Feature representation 2

Cosine similarity:
0.920
0.737

Motivation: Collecting Rich Training Data?

❖ Problems:

- Collecting complete training data with various poses, expressions, and illuminations is time-consuming and expensive
- Difficulty in collecting training samples in many practical applications (e.g., terrorists, public security, person Re-ID)

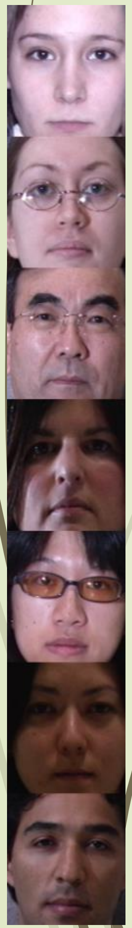


Scenarios: Low-Shot Face Recognition

- Scenario #1:
  - The face recognizer was trained on frontal faces (with neutral illumination and expression), and it cannot be retrained
  - Solution: **face normalization** (frontalization, deshading, and de-expression)
- Scenario #2:
  - Face recognizer is retrainable, but can only collected an anemic Training dataset
  - Solution: **face augmentation** (but how?)

Our solution: A unified **Do**main-**t**ransferred **F**ace **A**ugmentation **N**et (**DotFAN**) for **identity-preserving face augmentation & normalization**

Face Normalization (Deshading & Frontalization)

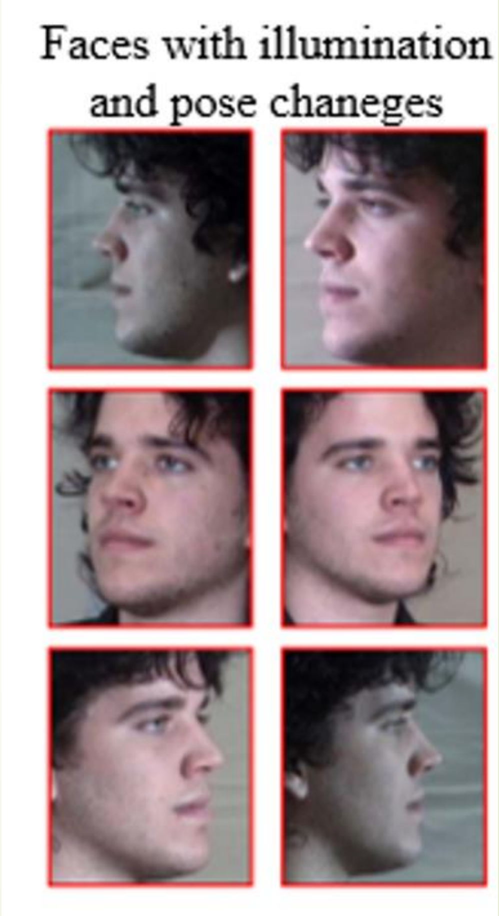


input

Deshading GAN

Deshaded

Face recognition



Face Augmentation

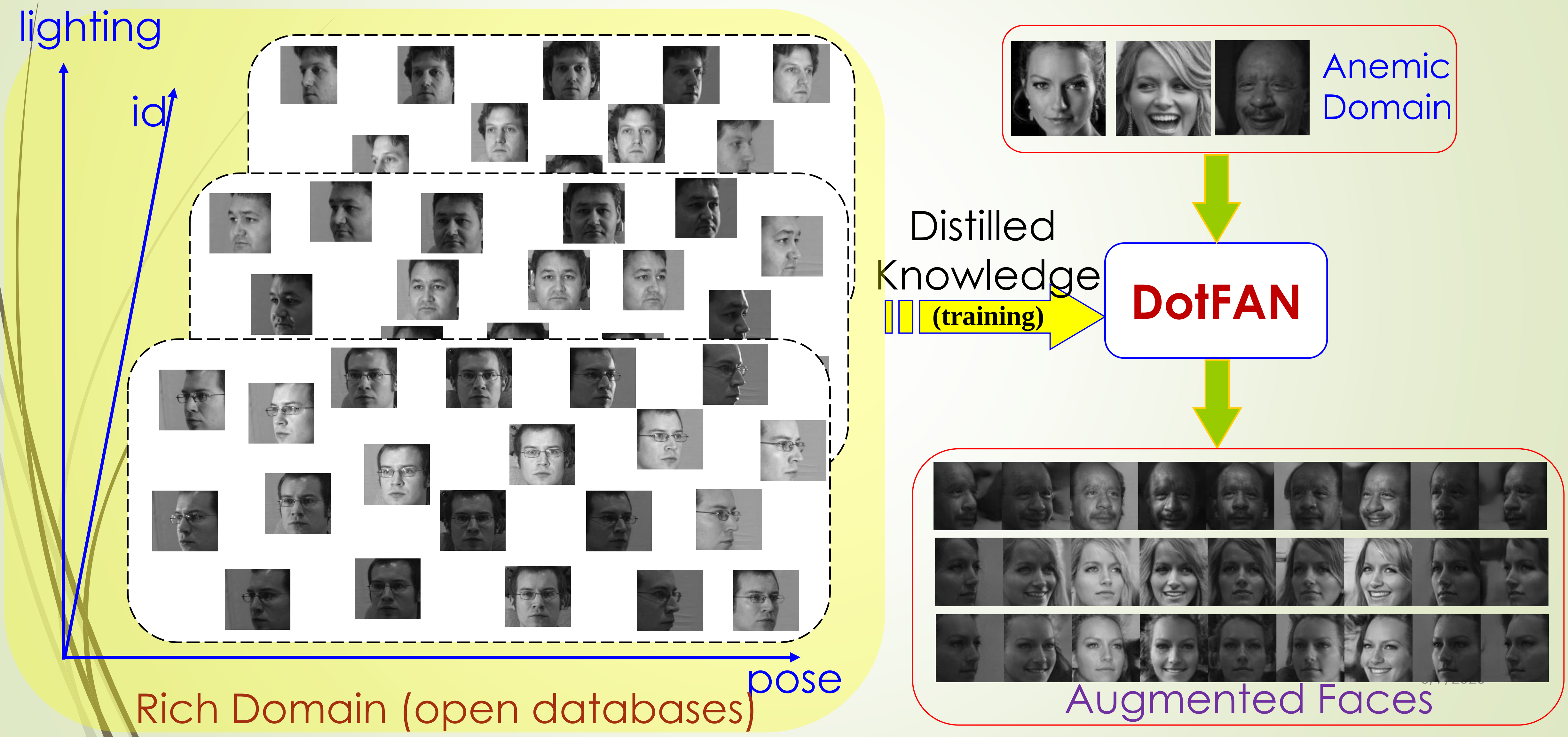
input

Synthesized

Face Augmentation

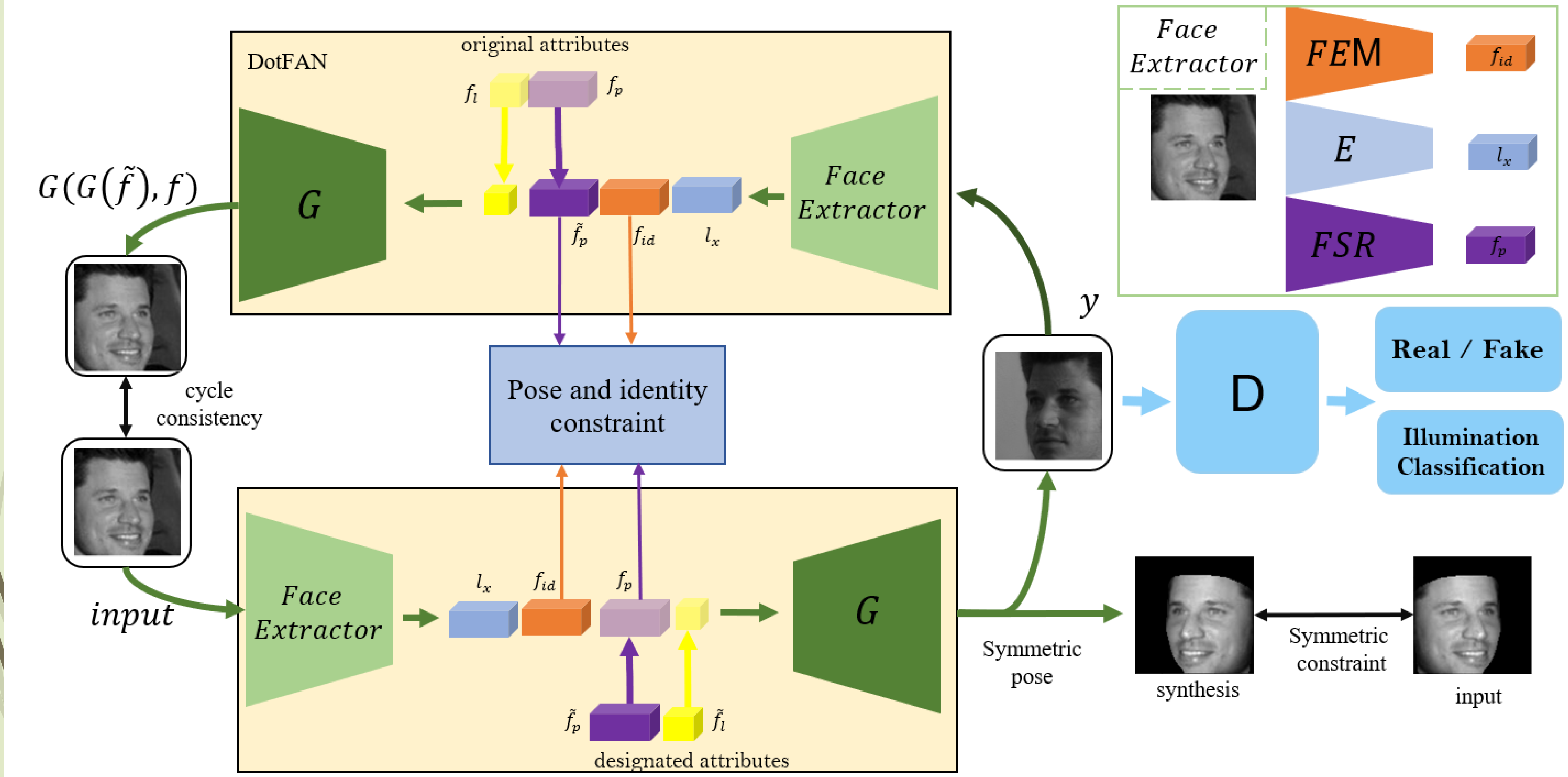
- Synthesizing faces of various poses and illumination patterns in unconstrained situations

Domain-transferred Face Augmentation Net (DotFAN)



DotFAN

FEM: Face-Expert Model
FSR: Face Shape Regressor



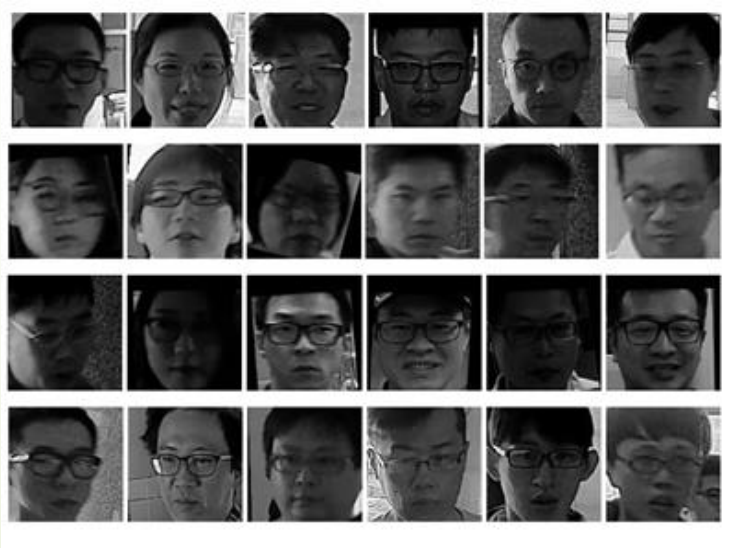
Experiments: Datasets

- Training datasets for FEM :
  - CASIA-WebFace
    - 494, 414 images from 10, 575 subjects
  - CelebA
    - 202, 599 images from 10, 177 subjects
    - Large-scale face attributes dataset
- Training datasets for FSR:
  - CASIA-Webface
  - CMU MultiPIE
    - Select 13 illumination ppatterns
    - 60 K faces

CMU Multi-PIE

Experiments: Datasets

- Testing Data
  - LFW
    - 13, 233 images of 5,749 identities
  - IJB-A
    - IJB-A has 25,808 images of 500 identities.
  - SurveilFace (LiteOn)
    - SurveilFace 1
      - 1,050 images of 73 identities
    - SurveilFace 2
      - 1,709 images of 78 identities



Face Normalization (Frontalization & Deshading)

- Verification accuracy on LFW

Methods	Verification Accuracy
3D	93.6±1.2
HPEN	96.3±0.8
FF-GAN	96.4±0.9
FaceID-GAN (CVPR18)	97.0±0.8
Proposed	**99.2±0.4**

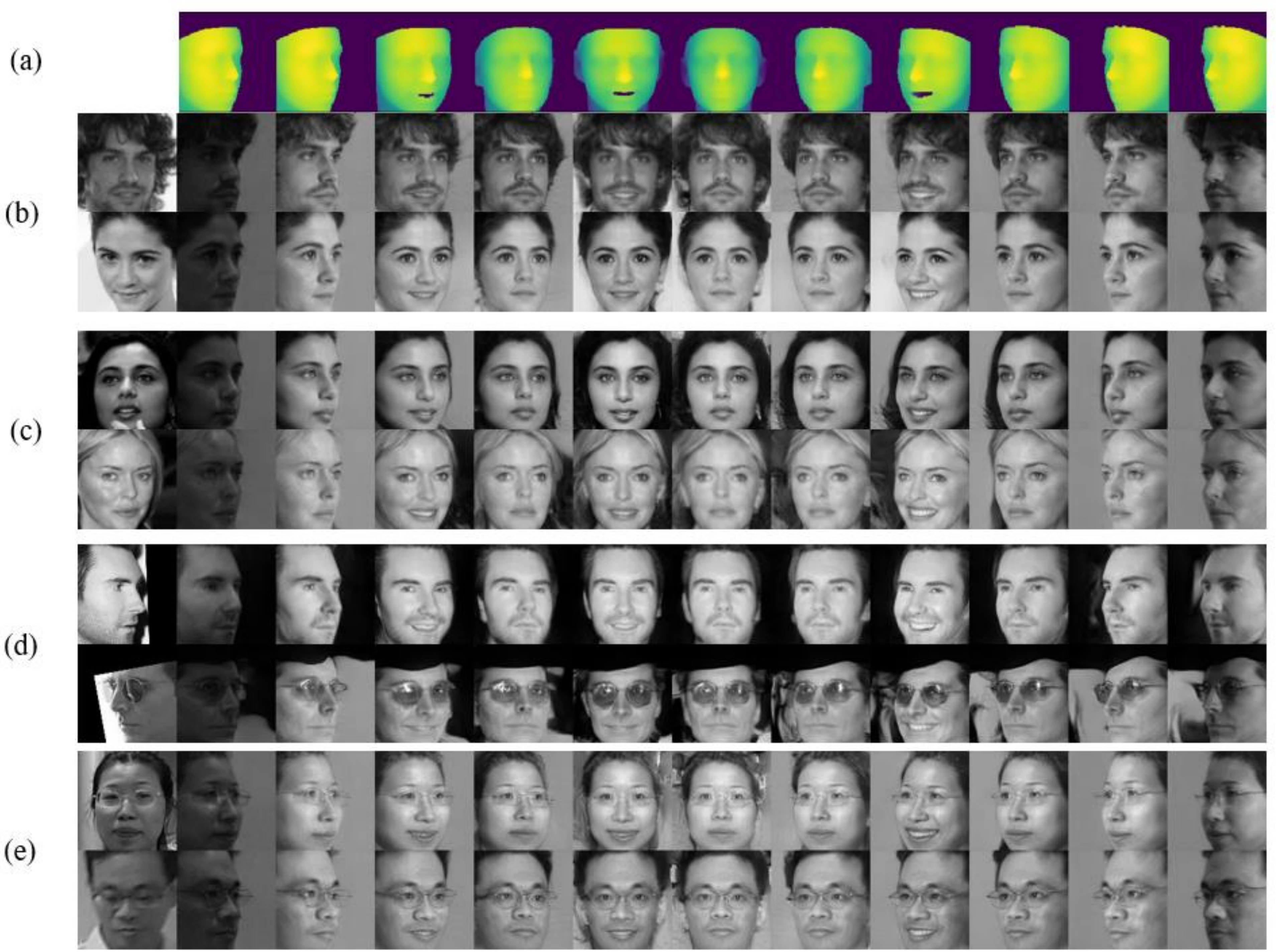
Face Normalization (Frontalization & Deshading)

- True-Acceptance-Rate (TAR) of verifications on IJB-A

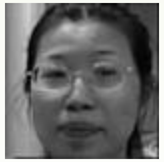
Methods	@0.01 FAR	@0.001 FAR
PAM	73.3±1.8	55.2±3.2
DCNN	78.7±4.3	-
DR-GAN	77.4±2.7	53.9±4.3
FF-GAN	85.2±1.0	66.3±3.3
FaceID-GAN [32]	87.6±1.1	69.2±2.7
Proposed	**93.7±0.5**	**89.3±1.0**

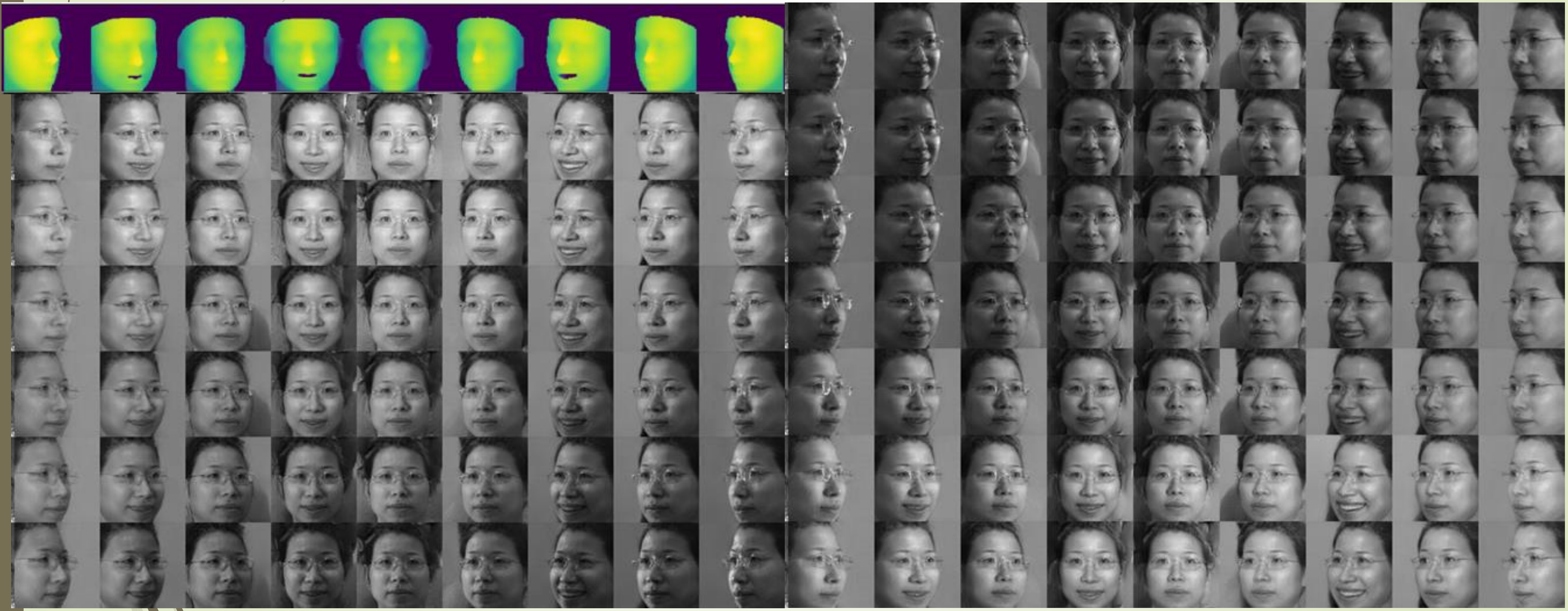
Augmentation

(a) 3D templates.
(b) CelebA
(c) LFW
(d) CFP
(e) SurveilFace



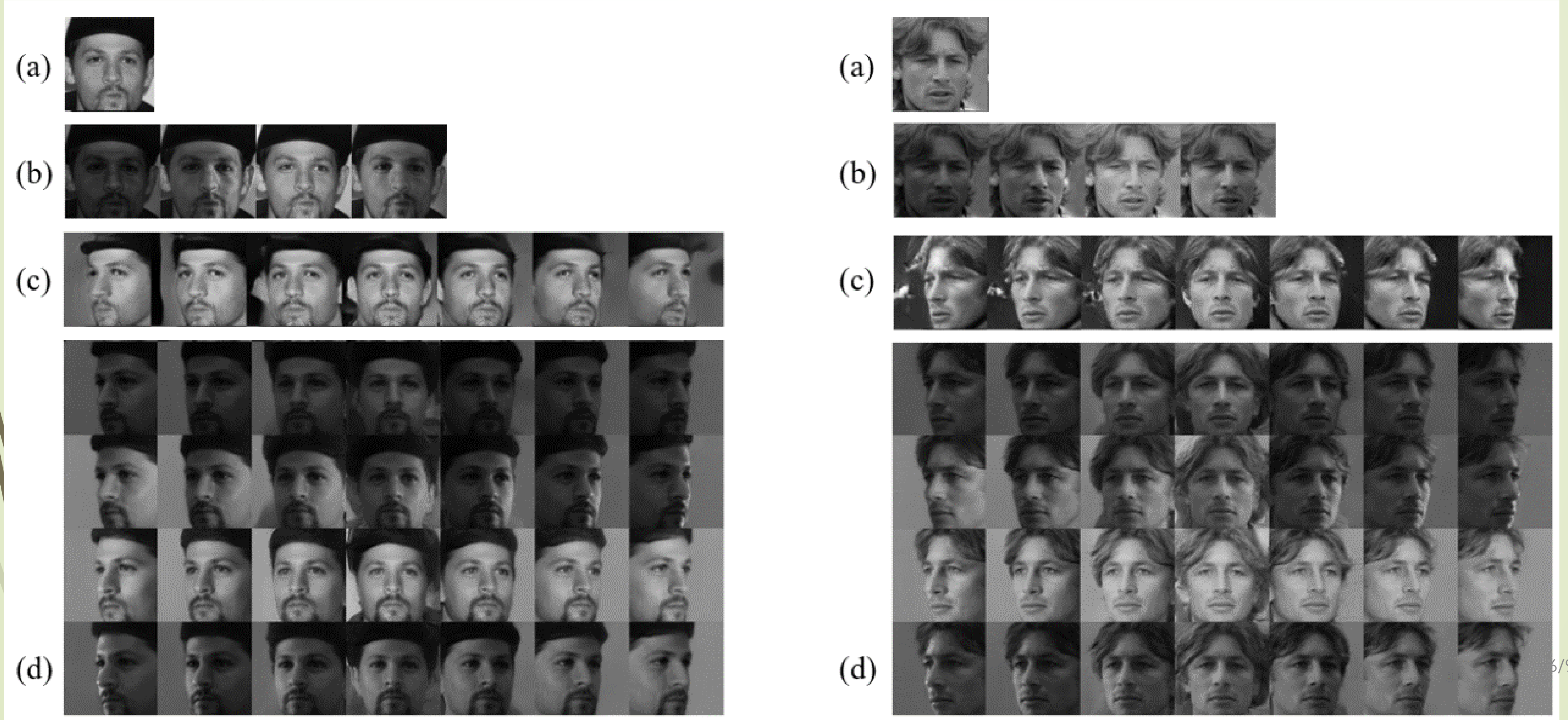
Augmentation: Full Illumination & Pose Codes





Augmentation: Comparison

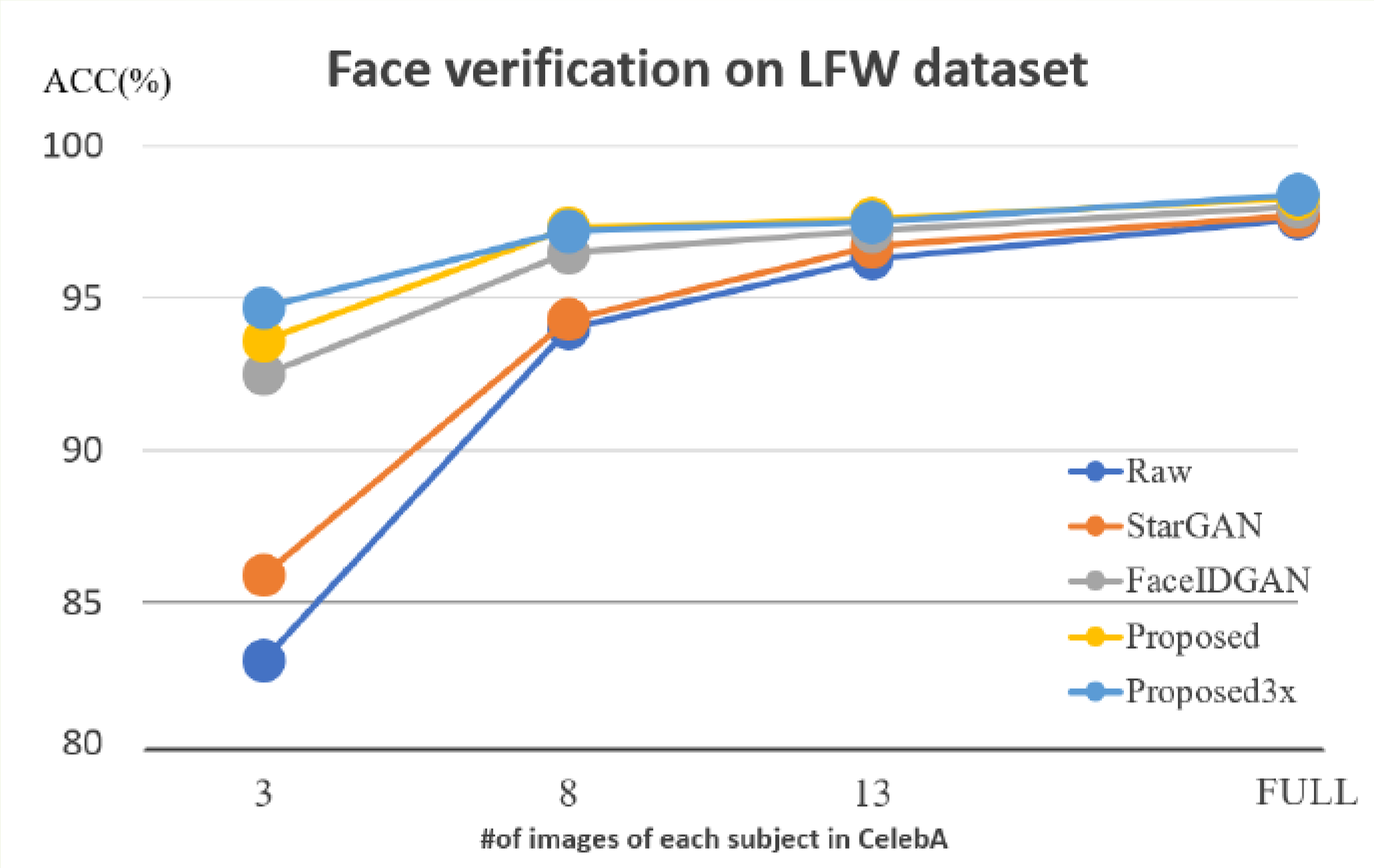
(a) input, (b) StarGAN, (c) FaceID-GAN, and (d) DotFAN (Proposed)



Face Recognition Performance Comparison

Method	LFW		SurveilFace-1			SurveilFace-2		
	ACC	AUC	@FAR=0.001	@FAR=0.01	AUC	@FAR=0.001	@FAR=0.01	AUC
(a) **Sub-CelebA(3)** (totally 30, 120 images)								
RAW	83.1	90.2	20.5	34.4	83.2	18.0	33.3	84.8
StarGAN	85.9	92.5	25.1	39.6	87.5	27.4	46.7	91.4
FaceID-GAN	92.5	97.6	34.6	53.5	92.8	32.3	54.0	94.3
Proposed 1x	93.6	98.1	35.7	56.2	93.6	34.7	57.8	95.0
Proposed 3x	94.7	98.7	36.8	58.3	94.6	36.5	60.8	95.6
(b) **Sub-CelebA(8)** (totally 75, 796 images)								
RAW	94.0	98.5	37.8	58.7	94.4	38.3	61.0	95.2
StarGAN	94.3	98.5	42.6	60.7	94.9	42.8	65.6	95.8
FaceID-GAN	96.5	99.3	48.1	65.6	96.0	45.7	67.9	96.8
Proposed 1x	97.3	99.5	53.2	71.2	97.0	49.1	72.2	97.2
Proposed 3x	97.2	99.5	53.2	68.9	96.9	47.3	70.0	97.1
(c) **Sub-CelebA(13)** (totally 116, 659 images)								
RAW	96.3	99.1	47.4	67.8	96.2	43.5	67.0	96.5
StarGAN	96.7	99.3	48.3	68.1	96.7	46.3	70.0	96.7
FaceID-GAN	97.2	99.5	53.3	71.3	97.0	50.2	72.3	97.4
Proposed 1x	97.6	99.6	56.2	75.1	97.7	50.4	73.9	97.7
Proposed 3x	97.5	99.7	56.7	75.5	97.7	53.9	72.2	97.8
(d) **CelebA (full CelebA dataset, 202, 599 images)**								
RAW	97.6	99.6	53.5	73.8	97.7	48.7	73.0	97.5
StarGAN	97.7	99.6	55.0	74.2	97.7	53.0	73.8	97.6
FaceID-GAN	98.0	99.7	57.6	76.4	98.1	54.1	76.5	98.0
Proposed 1x	98.3	99.8	62.4	80.9	98.4	57.1	76.7	98.1
Proposed 3x	98.4	99.7	61.4	78.9	98.2	54.7	77.8	98.0

Face Recognition Performance Comparison



Summary

- We propose GAN-based approaches to deal with the following difficult problems:
  - Face Normalization & Augmentation
    - Illumination variations
    - Pose variations
    - Both illumination & pose variations

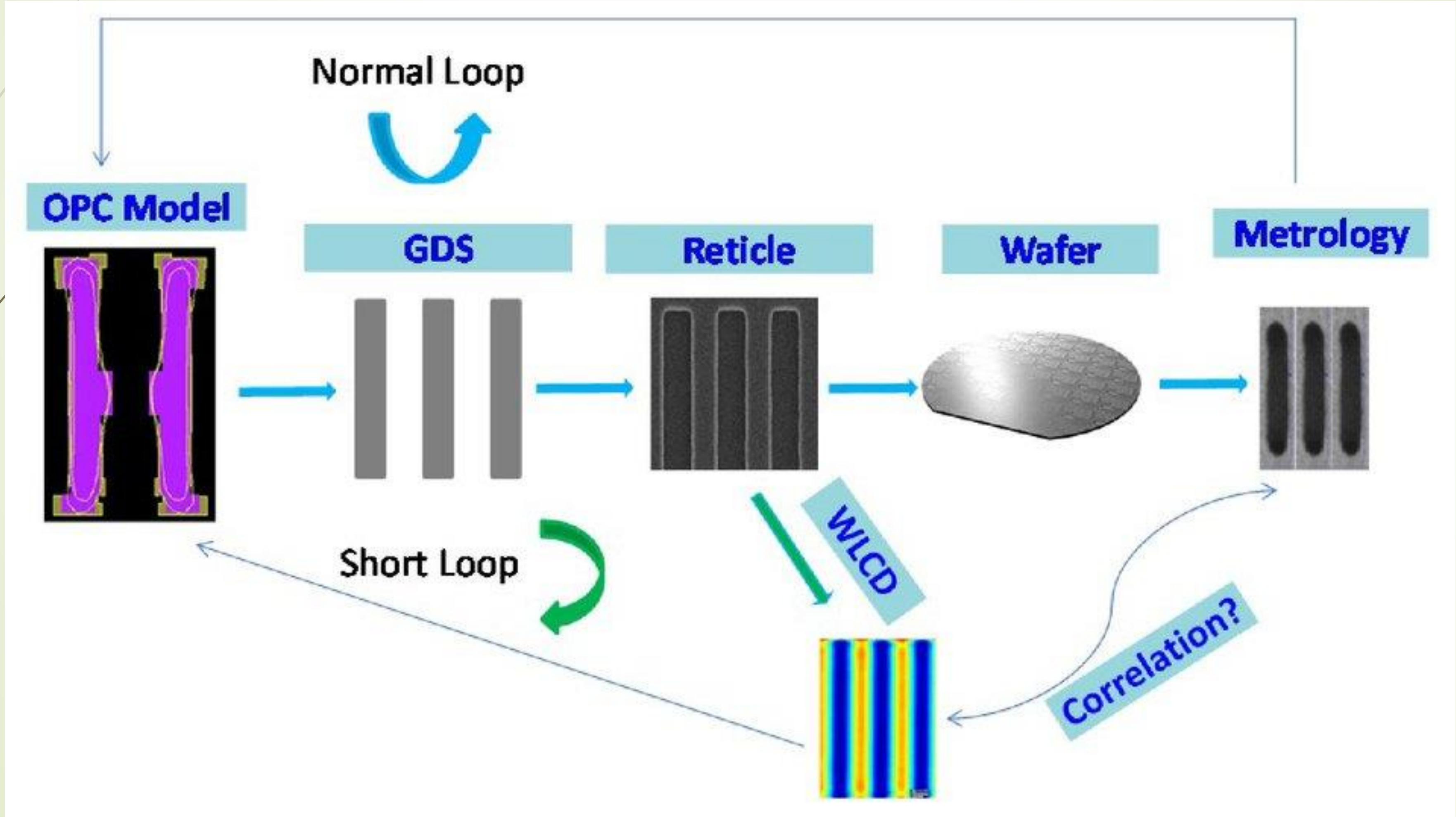
Industry Application
Data-Driven Prediction of IC Fabrication

Chia-Wen Lin (林嘉文), IEEE Fellow
Deputy Director, AI Technology Research Center
Professor, Department of Electrical Engineering
National Tsing Hua University, Taiwan
cwlin@ee.nthu.edu.tw

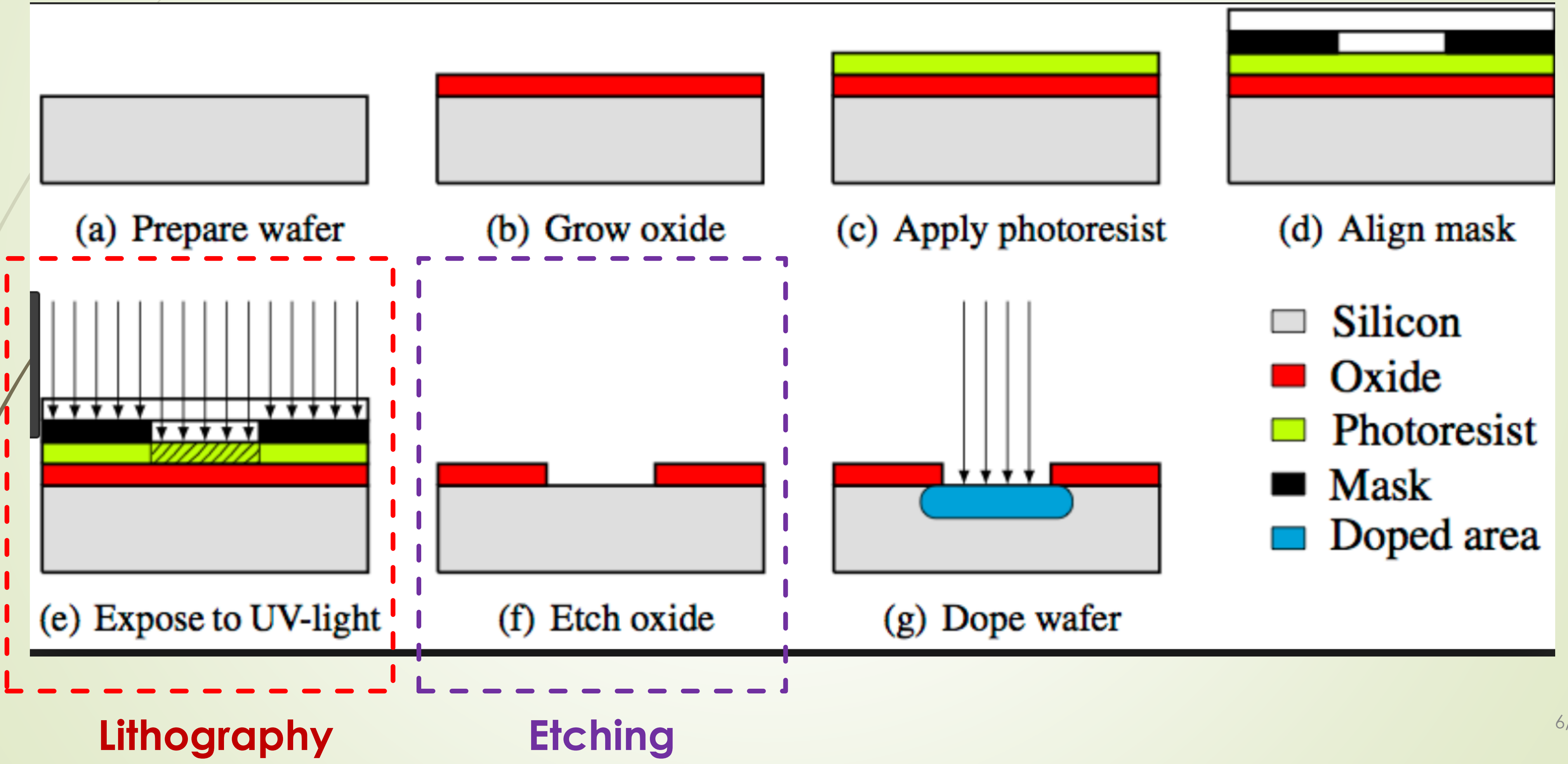
Outline

- Introduction
- LithoNet for Lithography Simulation
- OPCNet for Optical Proximity Correction
- Summary

IC Fabrication Process



IC Fabrication (Cont.)



Lithography

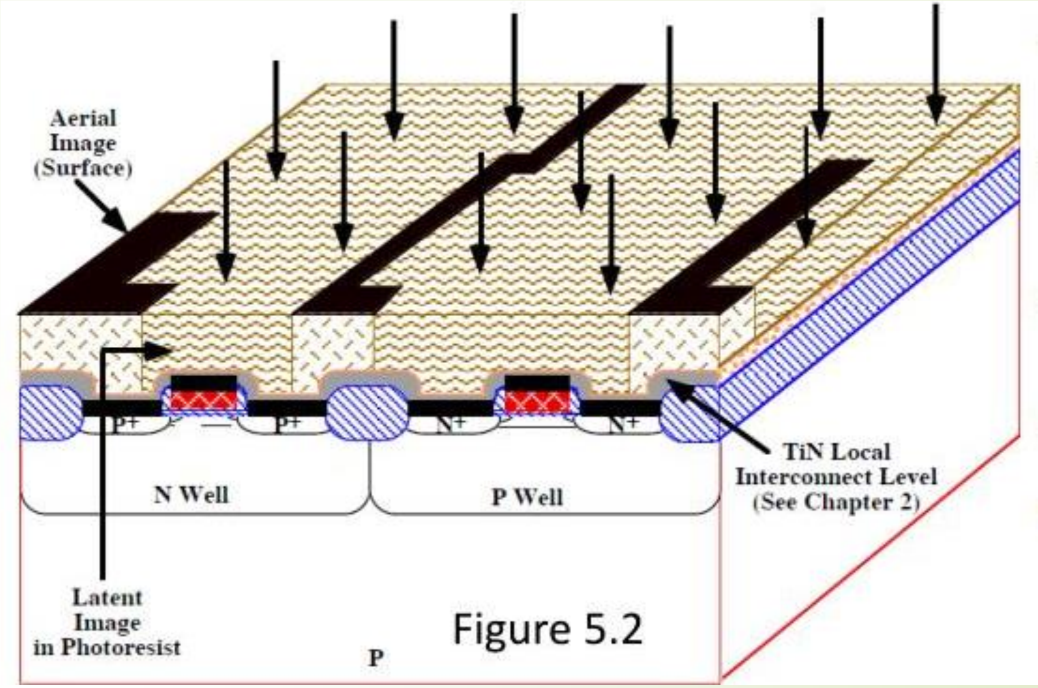
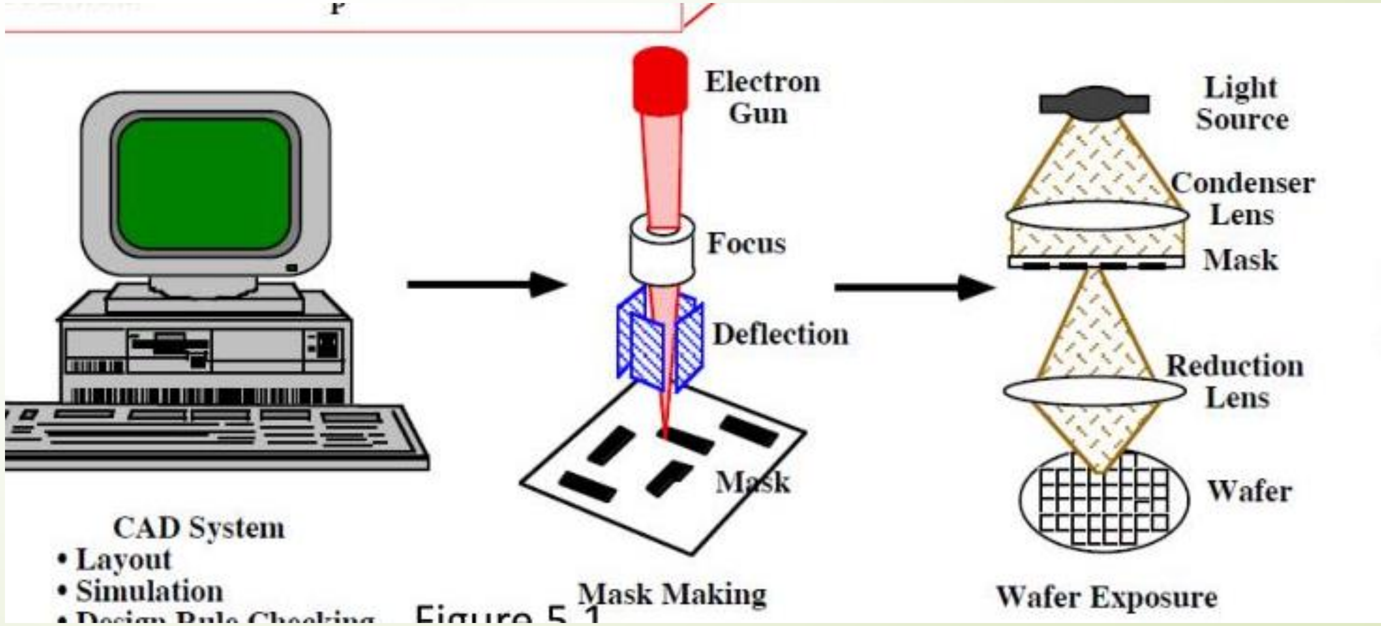
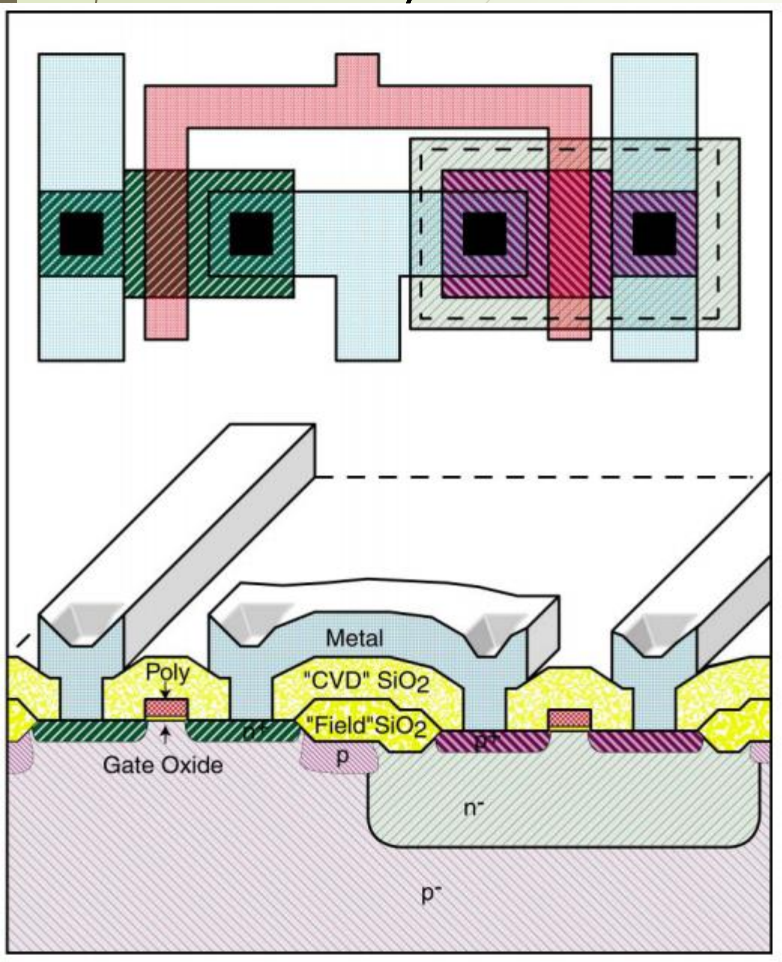


Figure 5.2

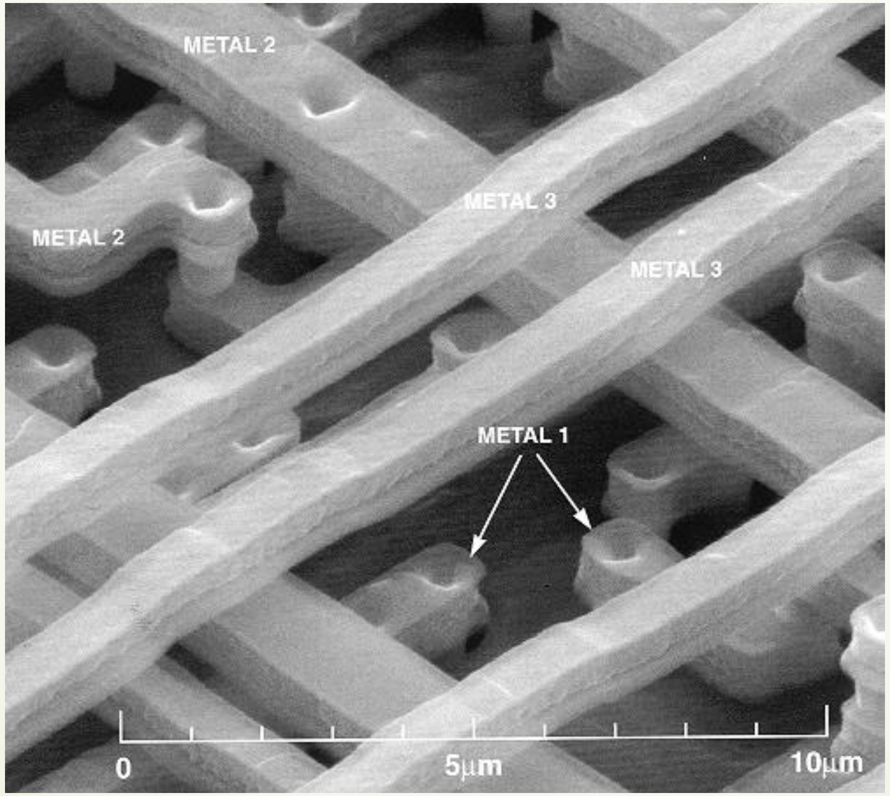


IC Fabrication (Cont.)

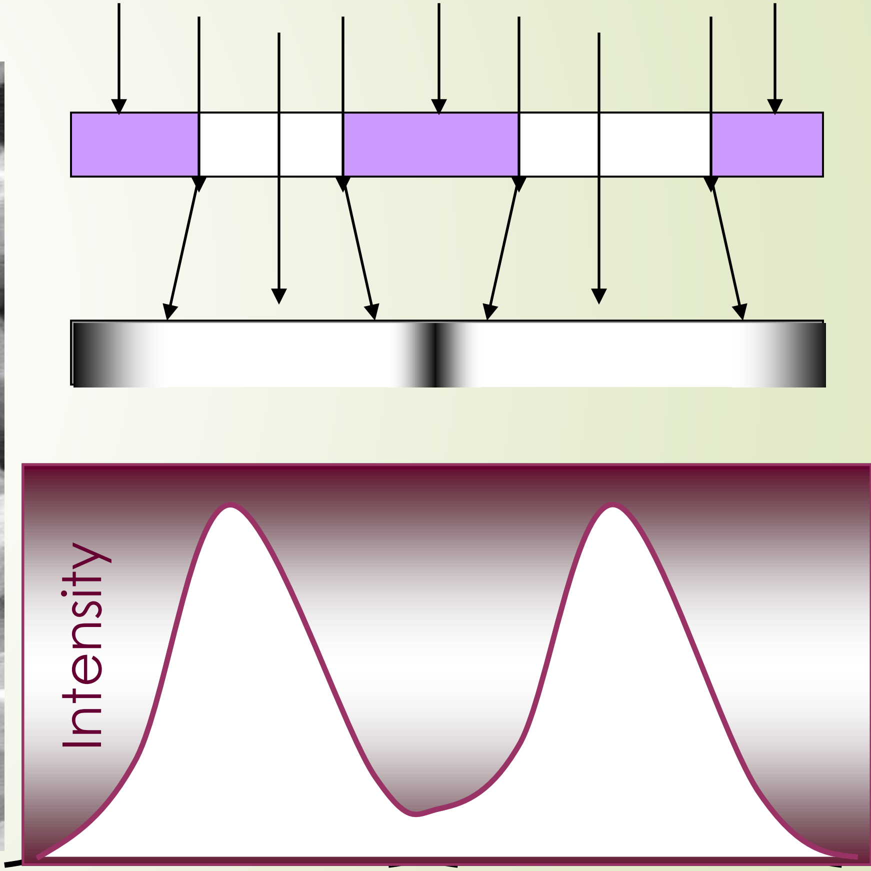
IC Layout



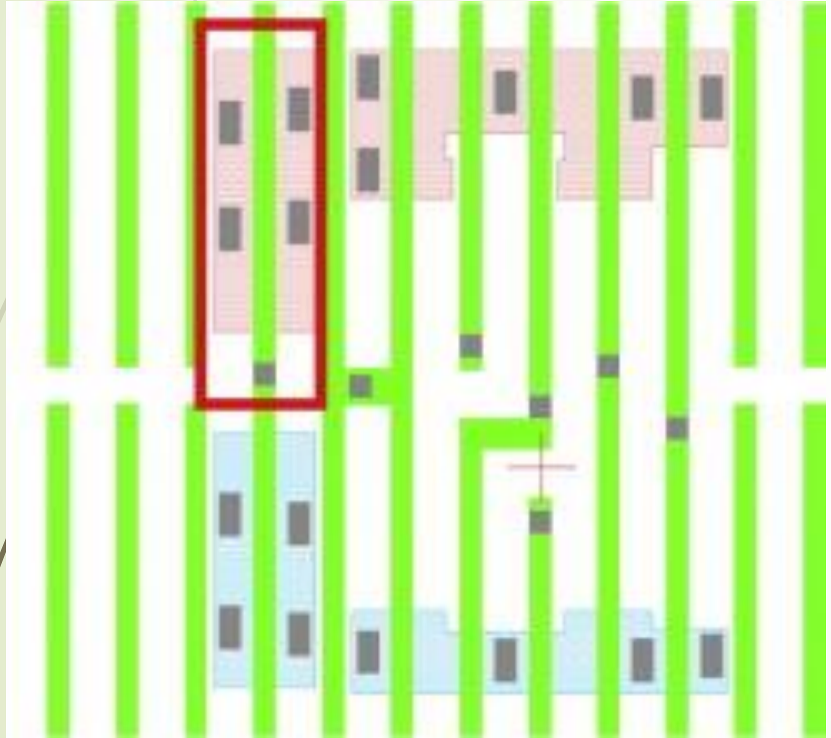
Fabricated Circuitry

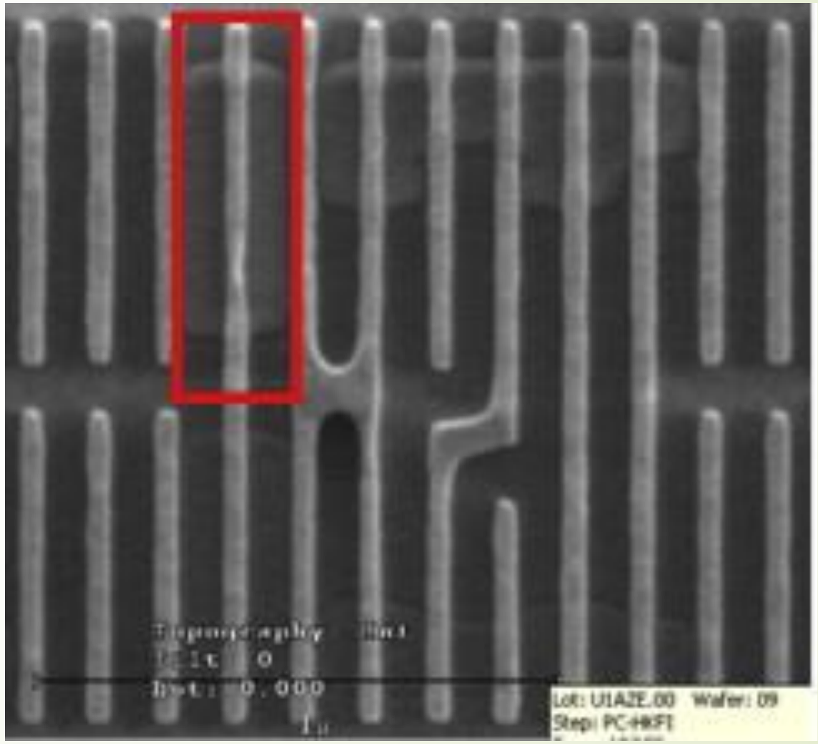


Masking & Lithography



IC Layout vs. Actual Circuitry (SEM Image)

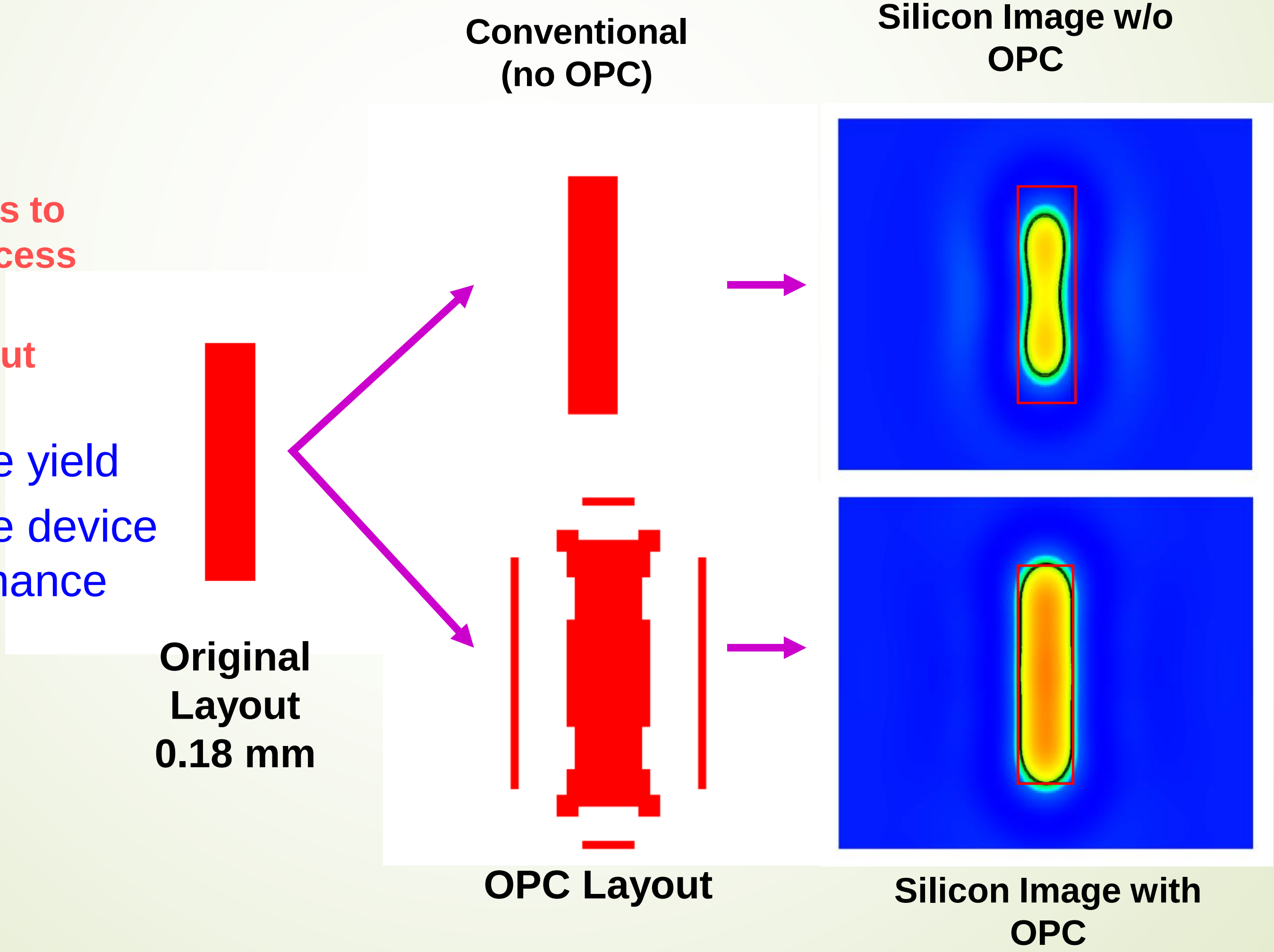




Lithography+Etching can be formulated as a nonlinear shape deformation process

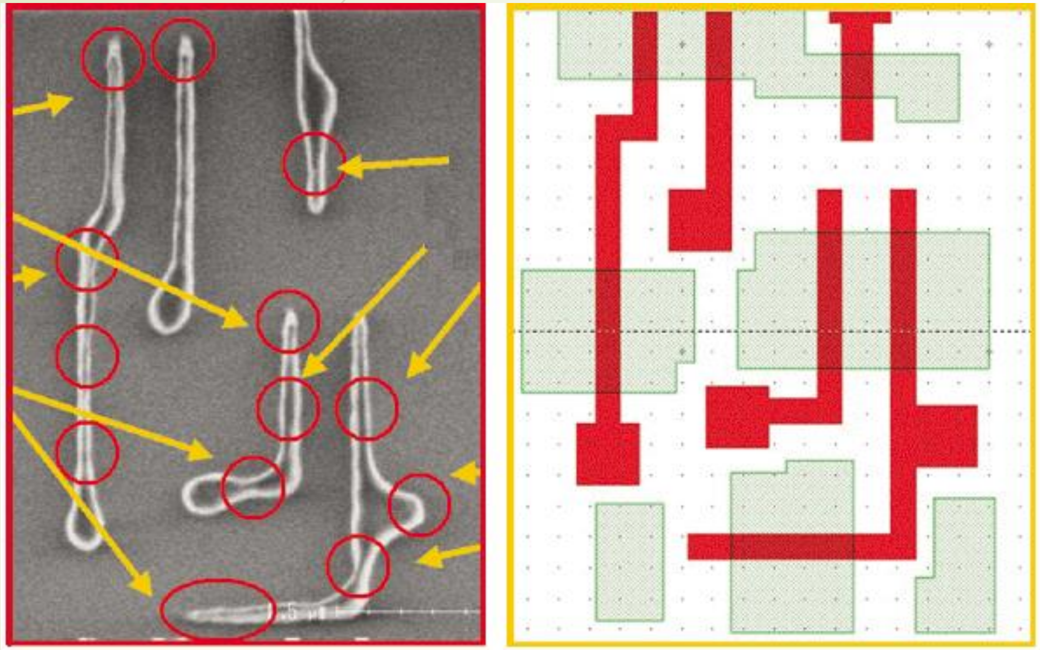
Optical Proximity Correction (OPC)

- Corrective modifications to improve process control
- As post-layout process
  - ➢ improve yield
  - ➢ improve device performance

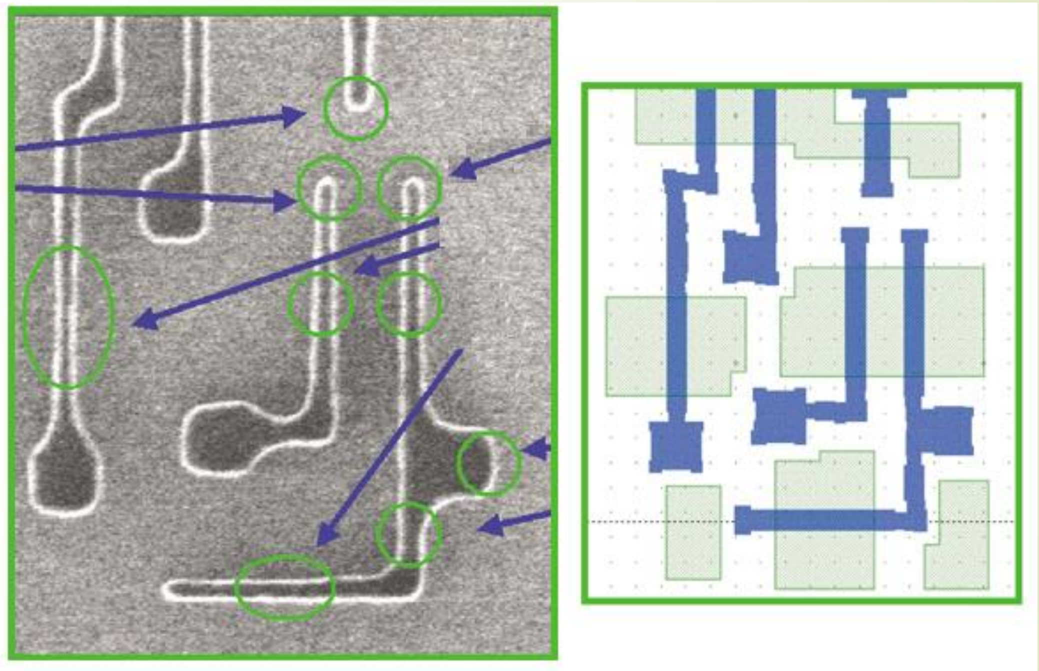


Optical Proximity Correction (OPC)

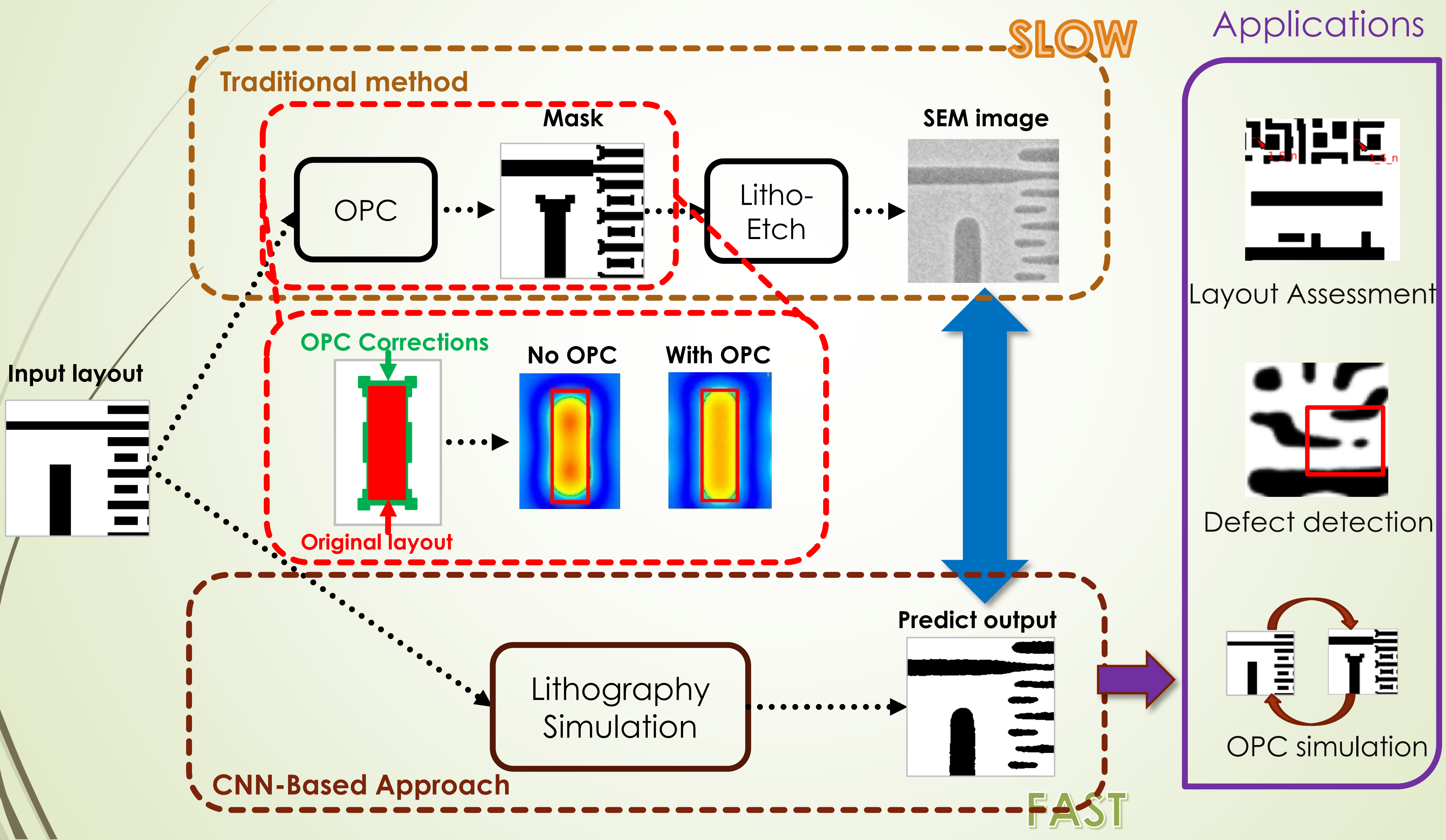
without OPC



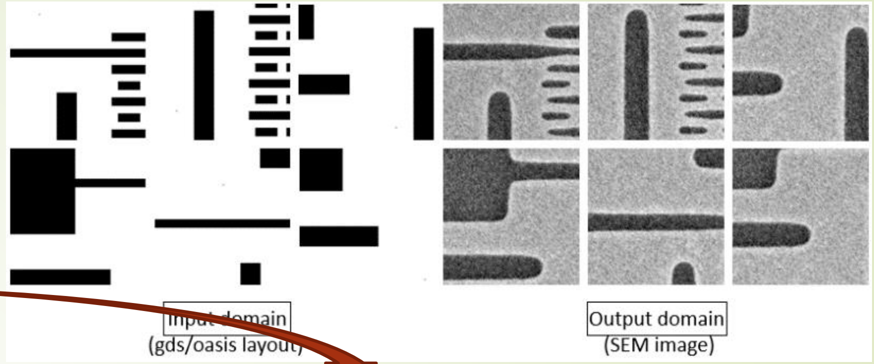
with OPC



IC Design for Manufacturability (DfM)

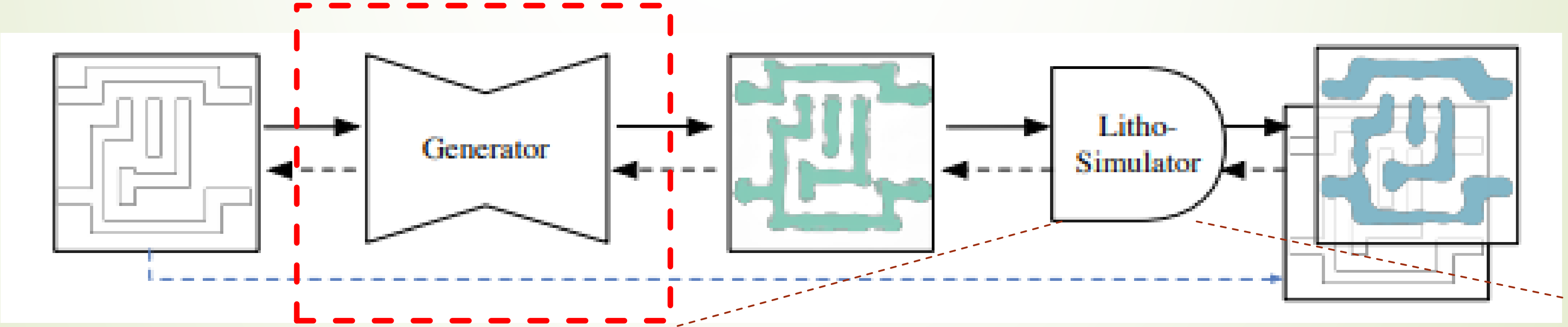


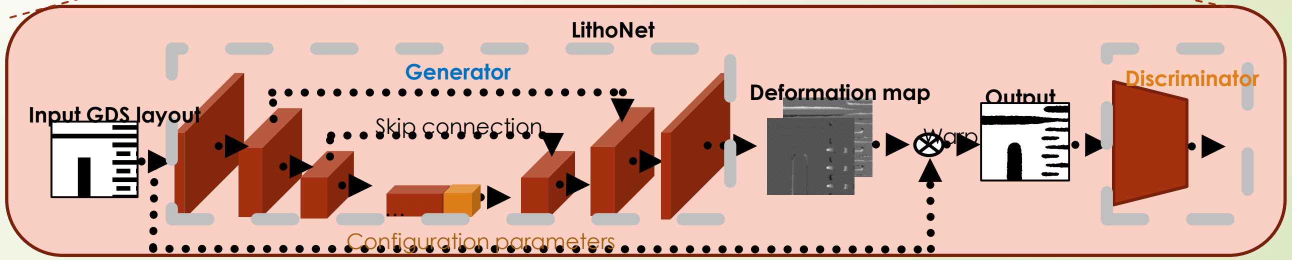
Research Goal



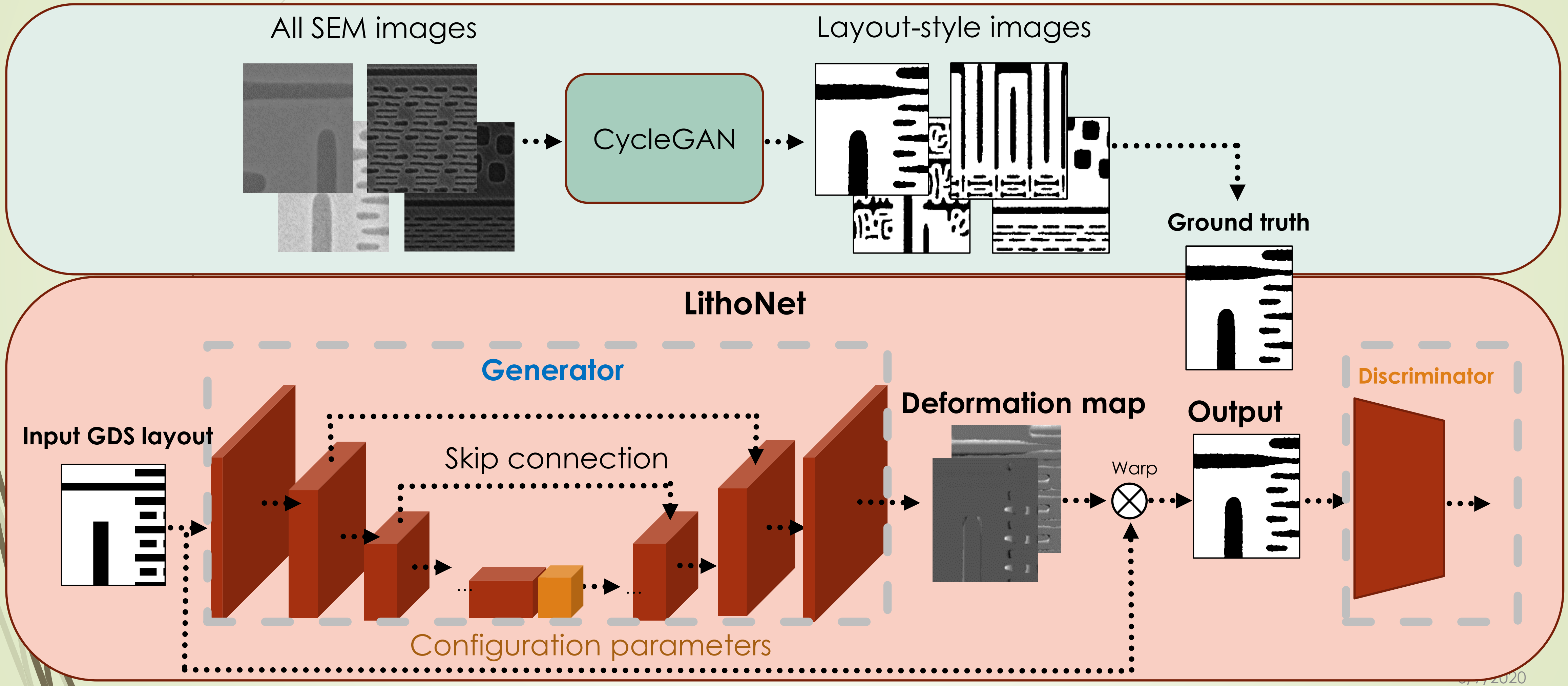
OPCNet

LithoNet

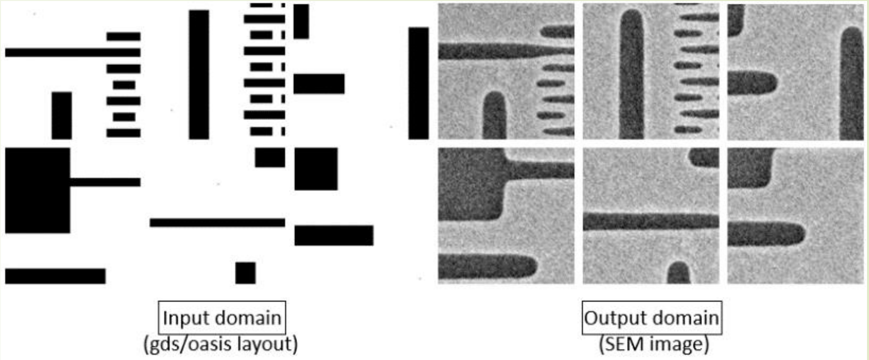




LithoNet for IC Fabrication Simulation



Experiment



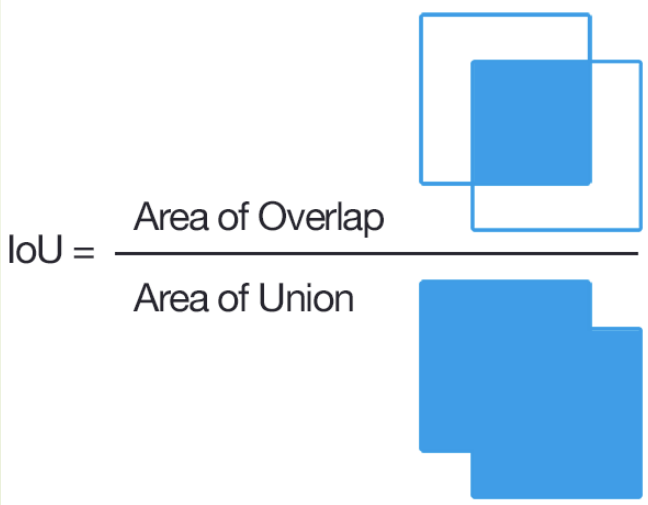
Datasets

1. UMC1
   Total **1,042 pairs** (942 pairs for training and 100 pairs for testing)
   Each pair contains a layout image and its SEM image

2. UMC2
   Total **7,483 pairs** (7,399 pairs for training and 84 pairs for testing)
   Seven different fabrication parameters (Ranging from -0.9 ~ +0.9)
   Each pair contains a layout image and its SEM image

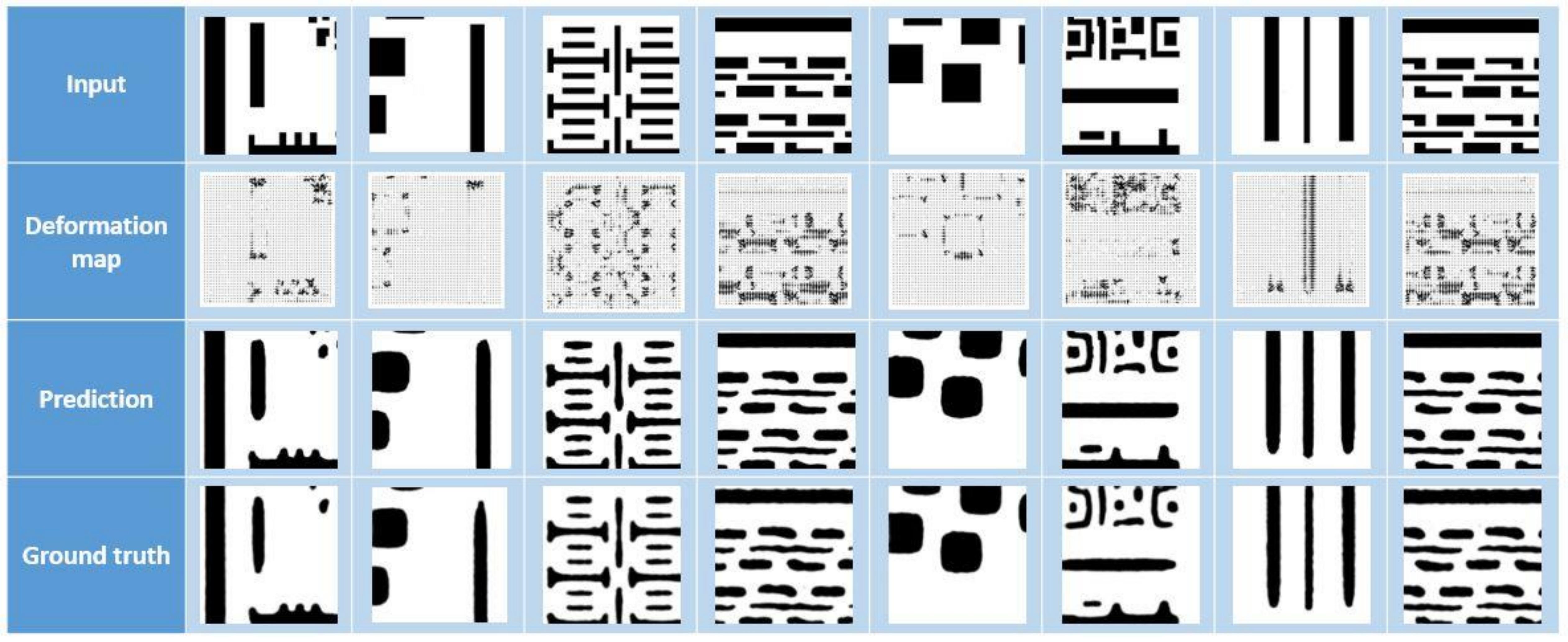
Experiments

- Metrics
  - Intersection Over Union (IoU)
  - SSIM
  - Per pixel error rate

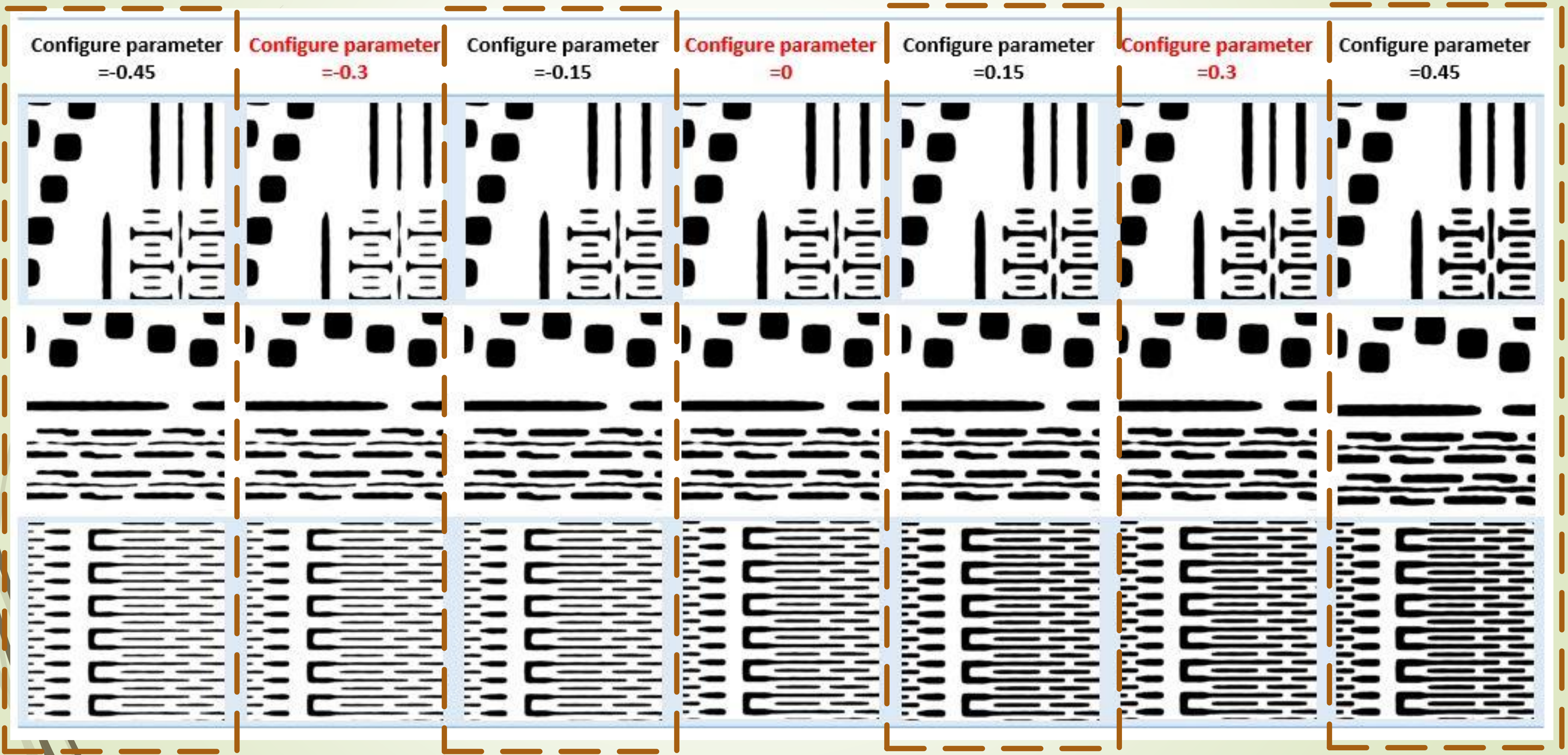


$$\mathrm{SSIM}(\mathbf{x},\mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}\text{。}$$

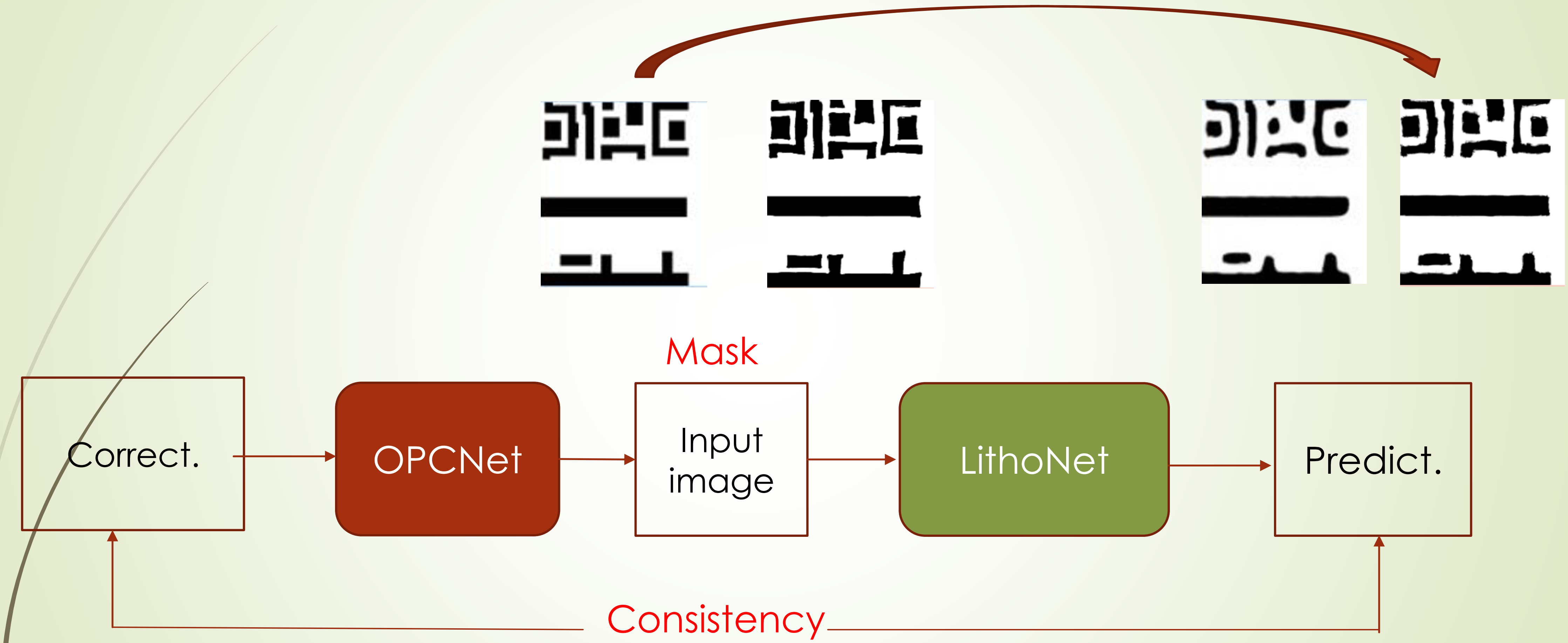
Evaluation of LithoNet



Prediction with New IC Fab Parameters



OPCNet for Mask Correction



Evaluation of OPCNet

Summary

- We proposed **LithoNet** and **OPCNet** for DfM of IC-Fab achieve promising accuracy and achieve 200x faster than an optical model-based prediction
- **Challenge**: Interdisciplinary research with partners having different technical background (also fun!)
- **Impact:** Lead to a paradigm shift in CAD tools for DfM of IC-Fab? (IC-Fab now hires an OPC team to establish the simulation models)
- Next research topics:
  - Low-shot learning (data-driven + model-based approach?)
  - Quality metrics for assessing IC layout

Conclusion

- The success of deep learning usually relies on the availability of large-scale well-labelled training data.
- In practice, it is often difficult to collect sufficient amount of well-labelled training samples.
- Low-shot learning has good potential to resolve part of deep learning problems in practical applications.