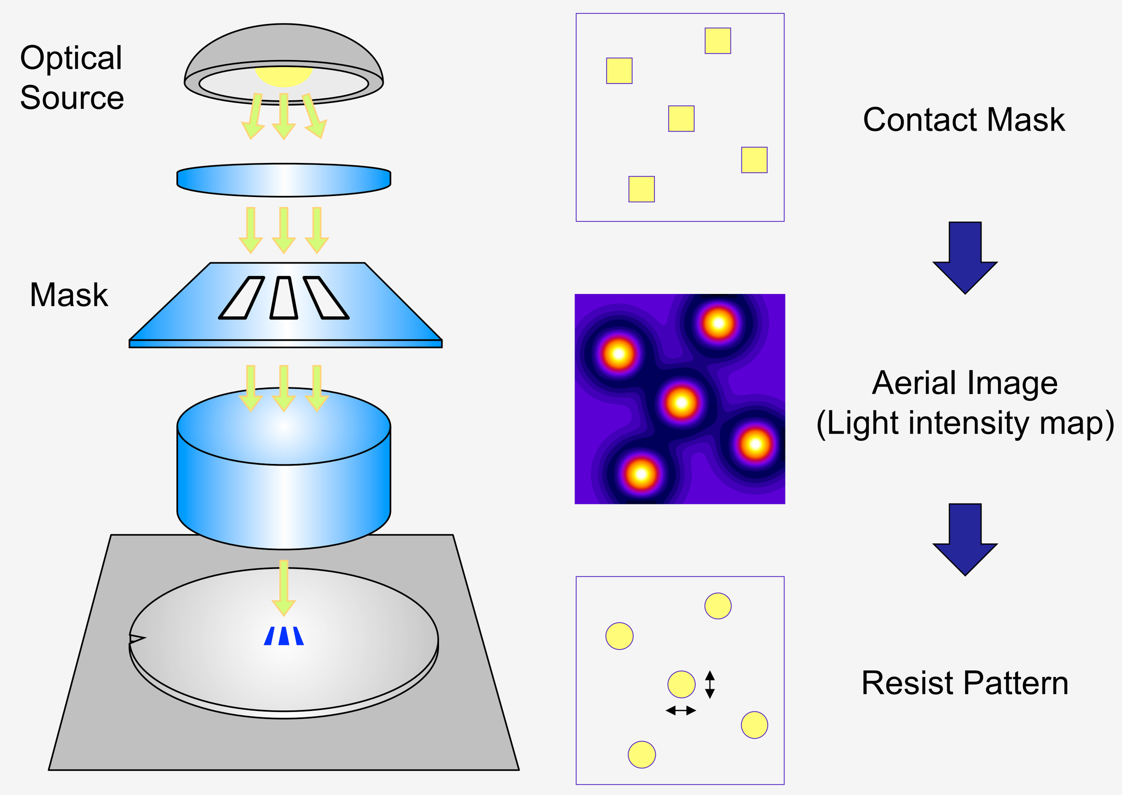
Data Efficient Lithography Modeling with Transfer Learning and Active Data Selection

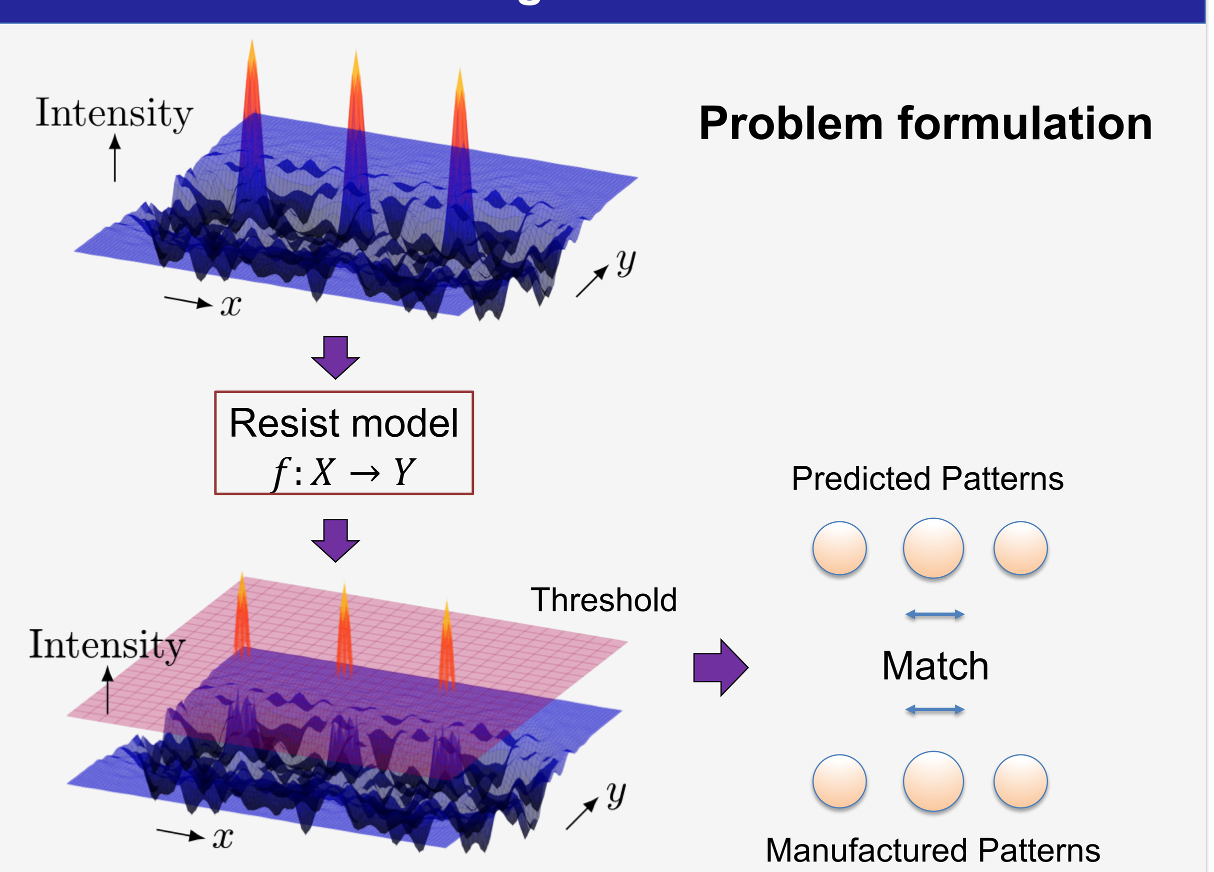
Yibo Lin[1], Meng Li[1], Yuki Watanabe[2], Taiki Kimura[2], Tetsuaki Matsunawa[2], Shigeki Nojima[2], David Z. Pan[1]

[1]University of Texas at Austin, [2]Toshiba Memory Corporation

Data Efficient Lithography Modeling



Modeling Photoresist

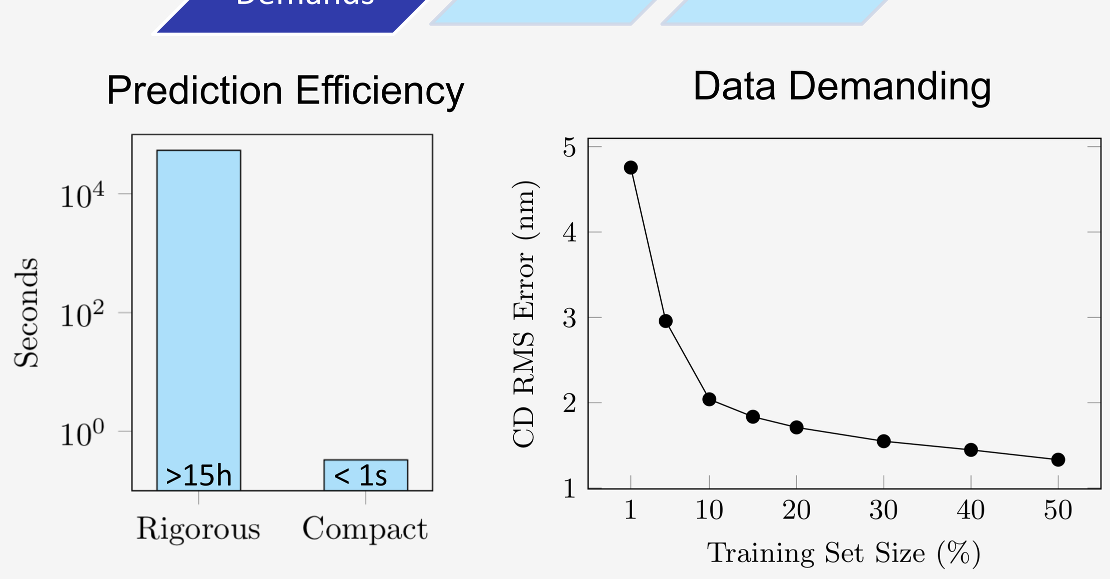


Challenges in Lithography Modeling

Rigorous simulation
- Physics-level simulation
- e.g., Synopsys Sentaurus Lithography

Compact model
- e.g., Mentor Graphics Calibre, machine learning models

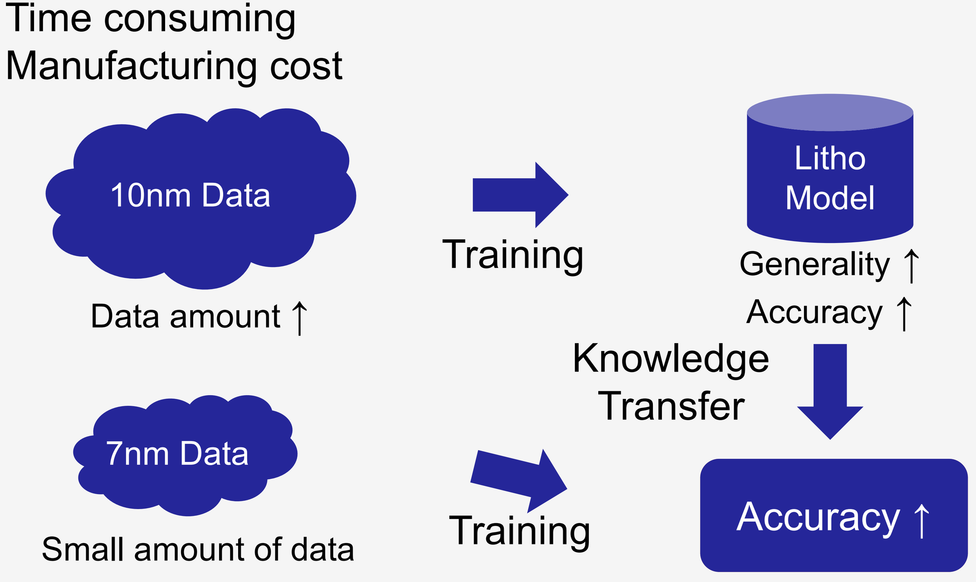
	Rigorous	Compact
Accuracy	High	Medium
Prediction Efficiency	Low	High
Data Demands	Medium	High



For 1K 2x2um² clips | Compact model

High Accuracy ➔ **big** training data

Expensive to prepare data
- Time consuming
- Manufacturing cost



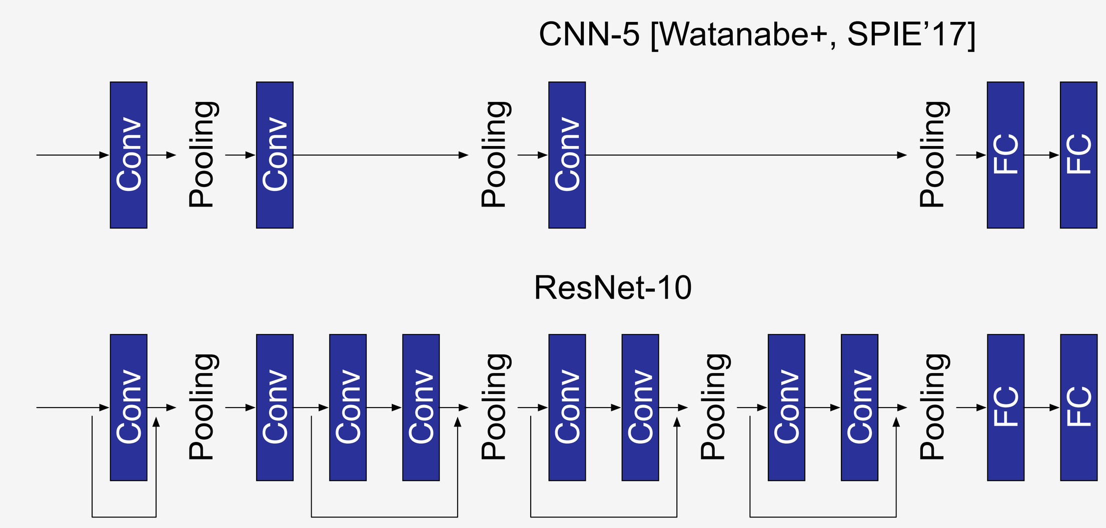
Deep Neural Networks for Lithography Modeling

Neural networks are getting **deeper** for higher accuracy

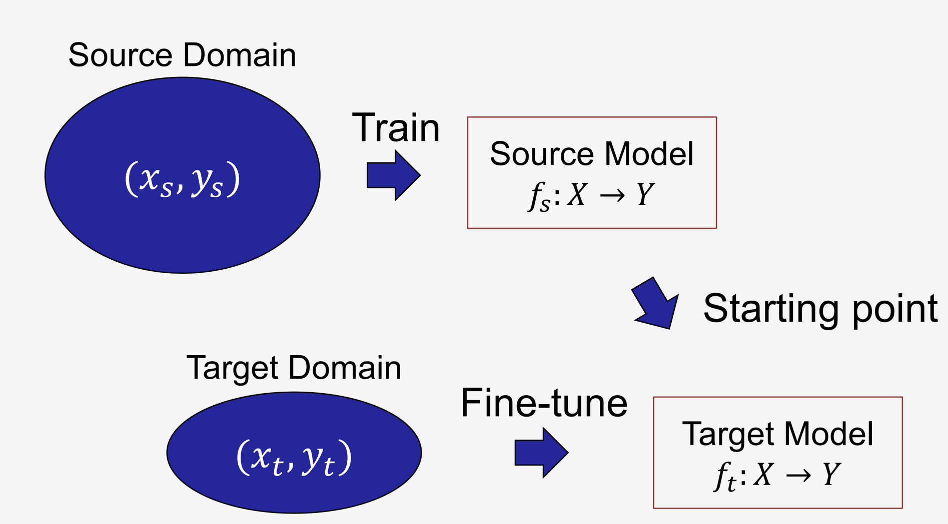
AlexNet-8, VGG-19, ResNet-101, ResNet-1202

> "You just keep on adding layers, until the test error doesn't improve anymore."
> – Yoshua Bengio

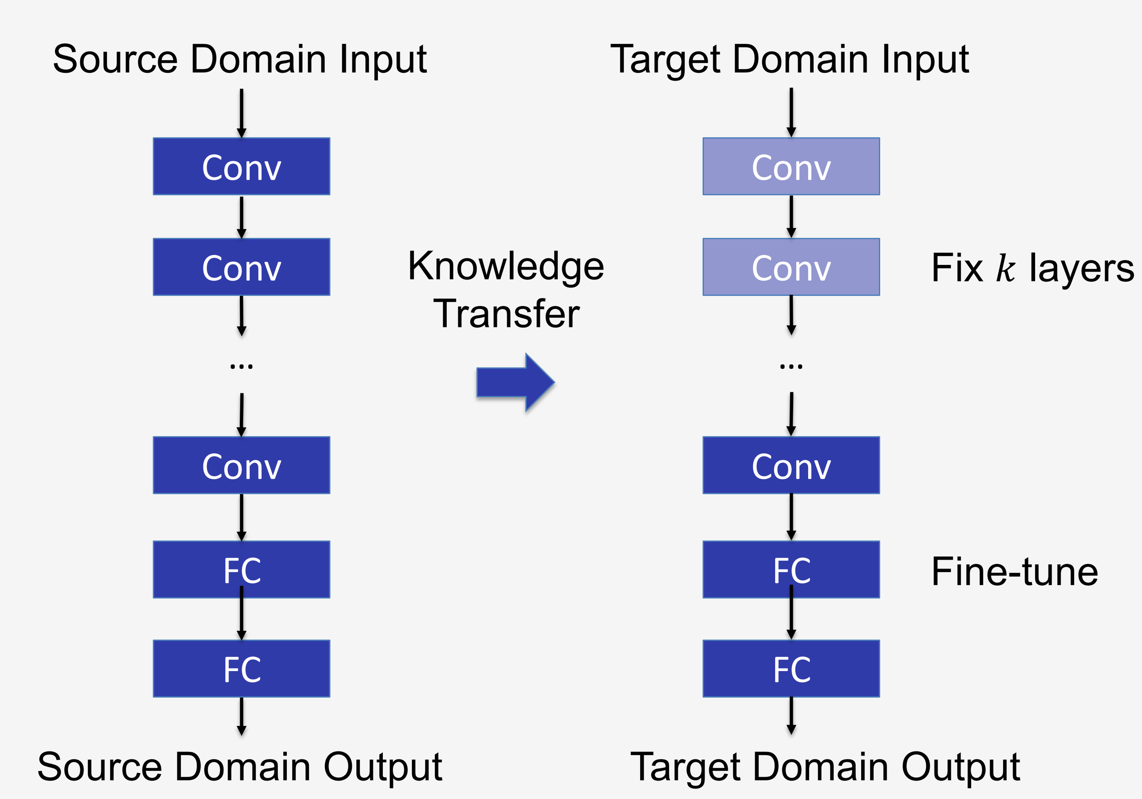
- Extend 5-layer CNN to 10-layer ResNet
- Solve gradient vanishing with shortcut connections



Transfer Learning for Lithography Modeling

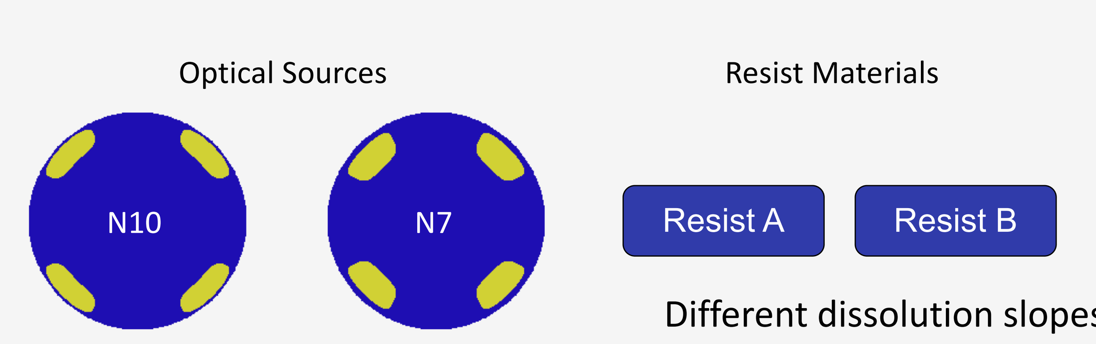


TF$_k$ scheme



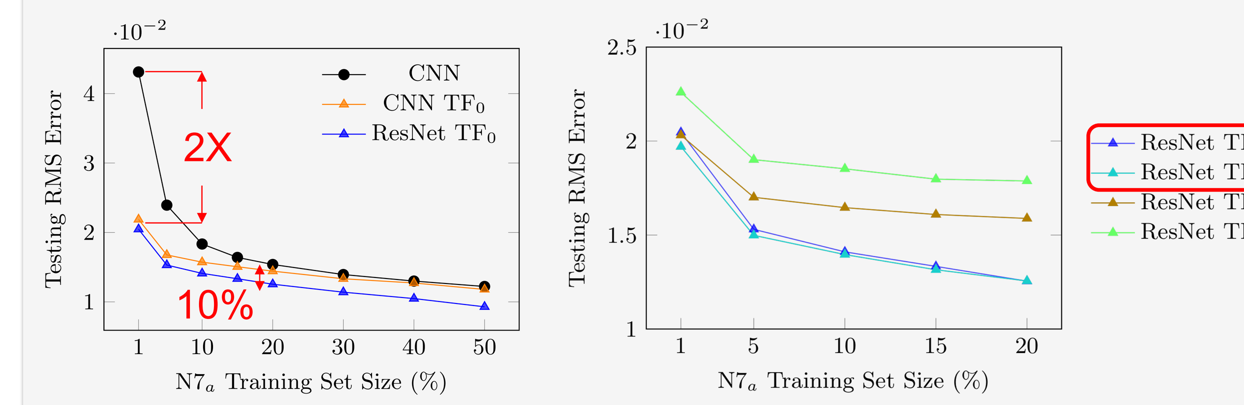
Technology Transition from N10 to N7

Contact Layer Design Rules [Liebmann, SPIE'15]	N10	N7
Patterning	LELE	LELELE
Pitch (nm)	64	45
Mask pitch (nm)	128	135
Litho-target (nm)	60	60

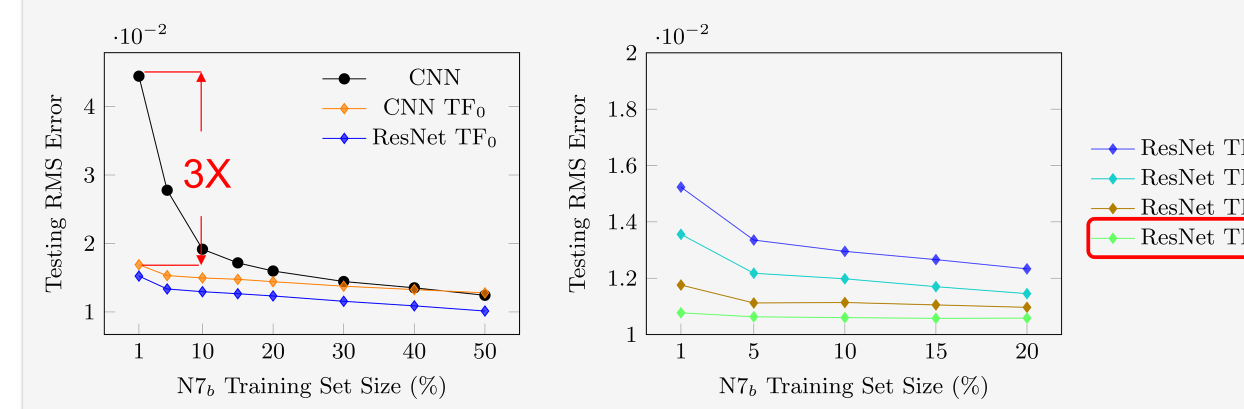


	N10	N7$_a$	N7$_b$
Design Rule	A	B	B
Optical Source	A	B	B
Resist Material	A	A	B

Case I: From N10 to N7



Case II: From N7$_a$ to N7$_b$



	Case I	Case II
Knowledge Transfer	N10➔N7$_a$	N7$_a$➔N7$_b$
Dataset Similarity	Medium	High
Best k	0/4	8

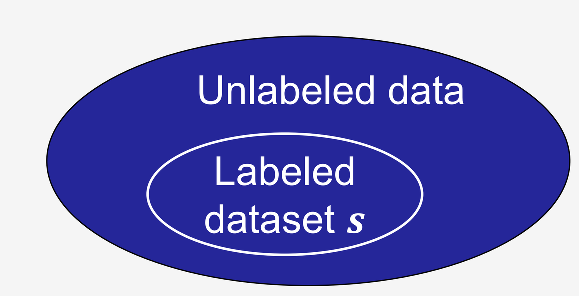
Active Learning with Clustering

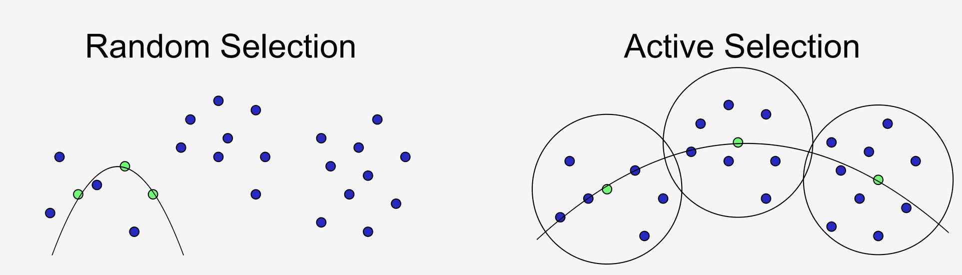
Target Domain Dataset D

$\forall (x, ?) \in D$

Making x is cheap

Querying ? is **expensive**





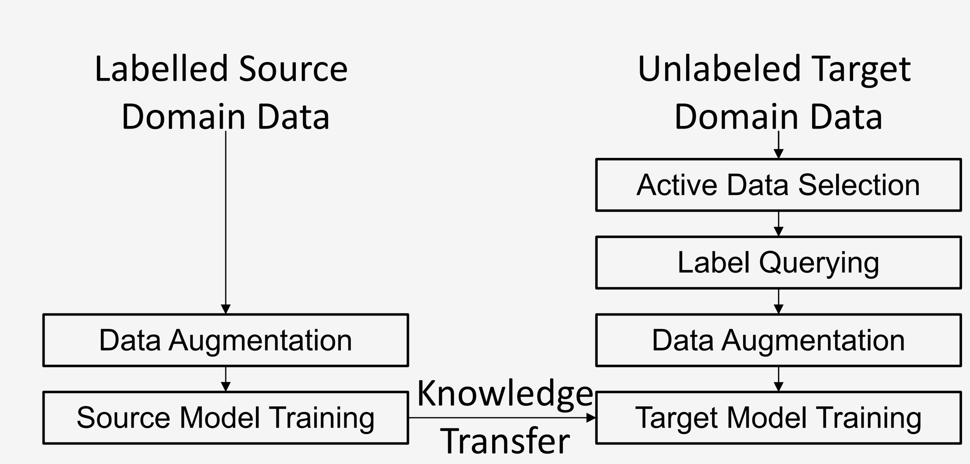
Intuition: close features have close labels

Theorem

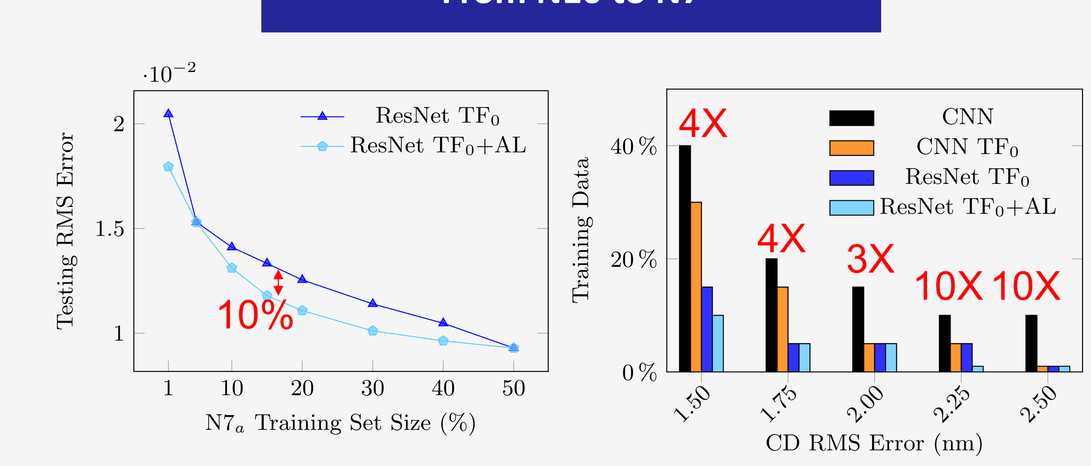
$$\frac{1}{n}\sum_{i \in D} Loss_i \leq \frac{C}{n}\sum_{j \in s}\sum_{i=1}^{k_j}\|x_i - x_j^c\| + \epsilon$$

K-Medoids Clustering
(A variation of K-Means Clustering)

Transfer & Active Learning Flow



From N10 to N7



- Improve data efficiency
- 3~10X reduction of training data
- Reduce turn-around time
- Increase model accuracy