Cover-source Mismatch in Deep Spatial Steganalysis

IWDW
2019

Outlines

Outlines

1. Motivation

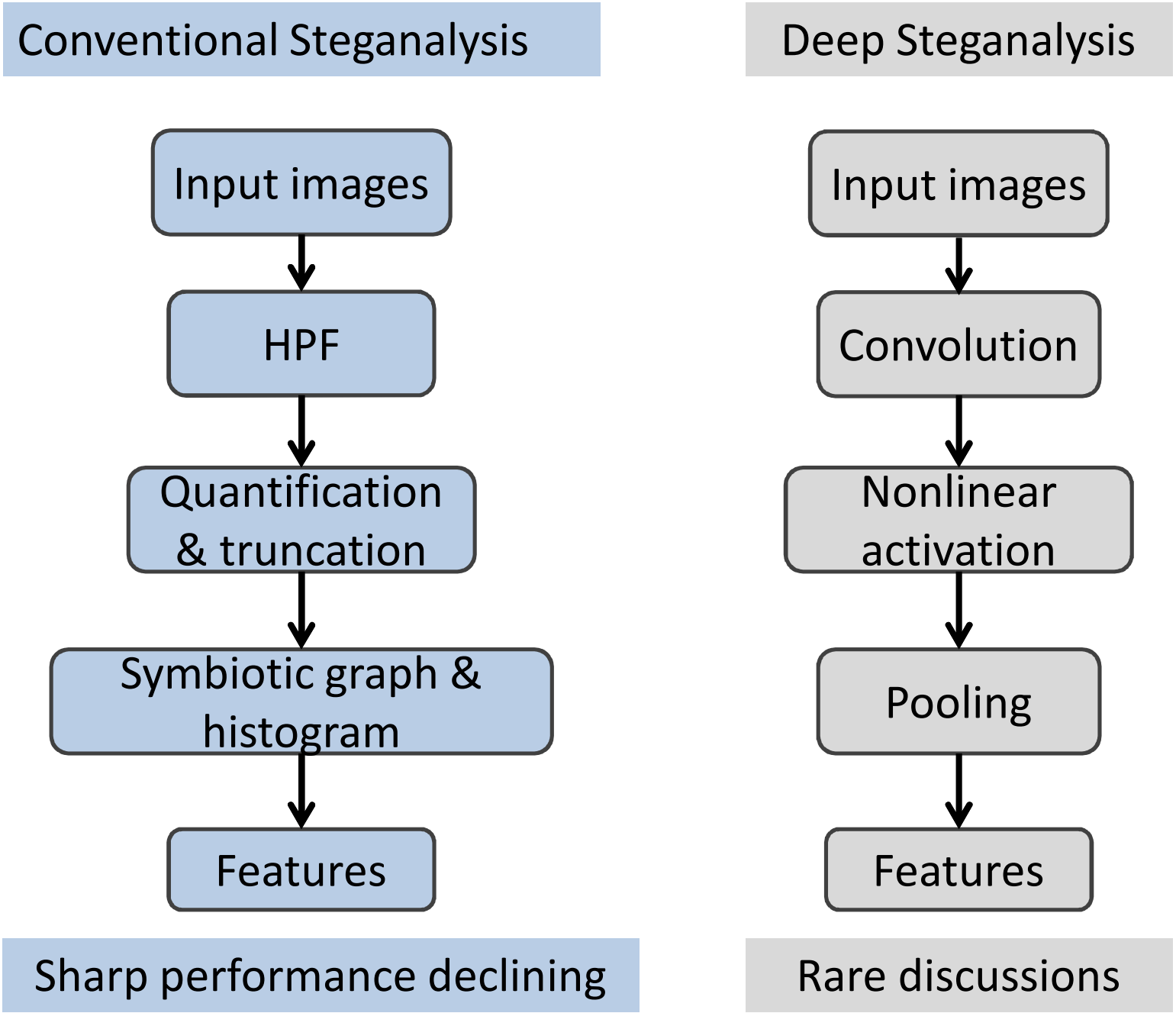
- **Cover-source Mismatch**




Training set




Testing set


Conventional Steganalysis
Input images
HPF
Quantification & truncation
Symbiotic graph & histogram
Features
Sharp performance declining
Deep Steganalysis
Input images
Convolution
Nonlinear activation
Pooling
Features
Rare discussions

Outlines

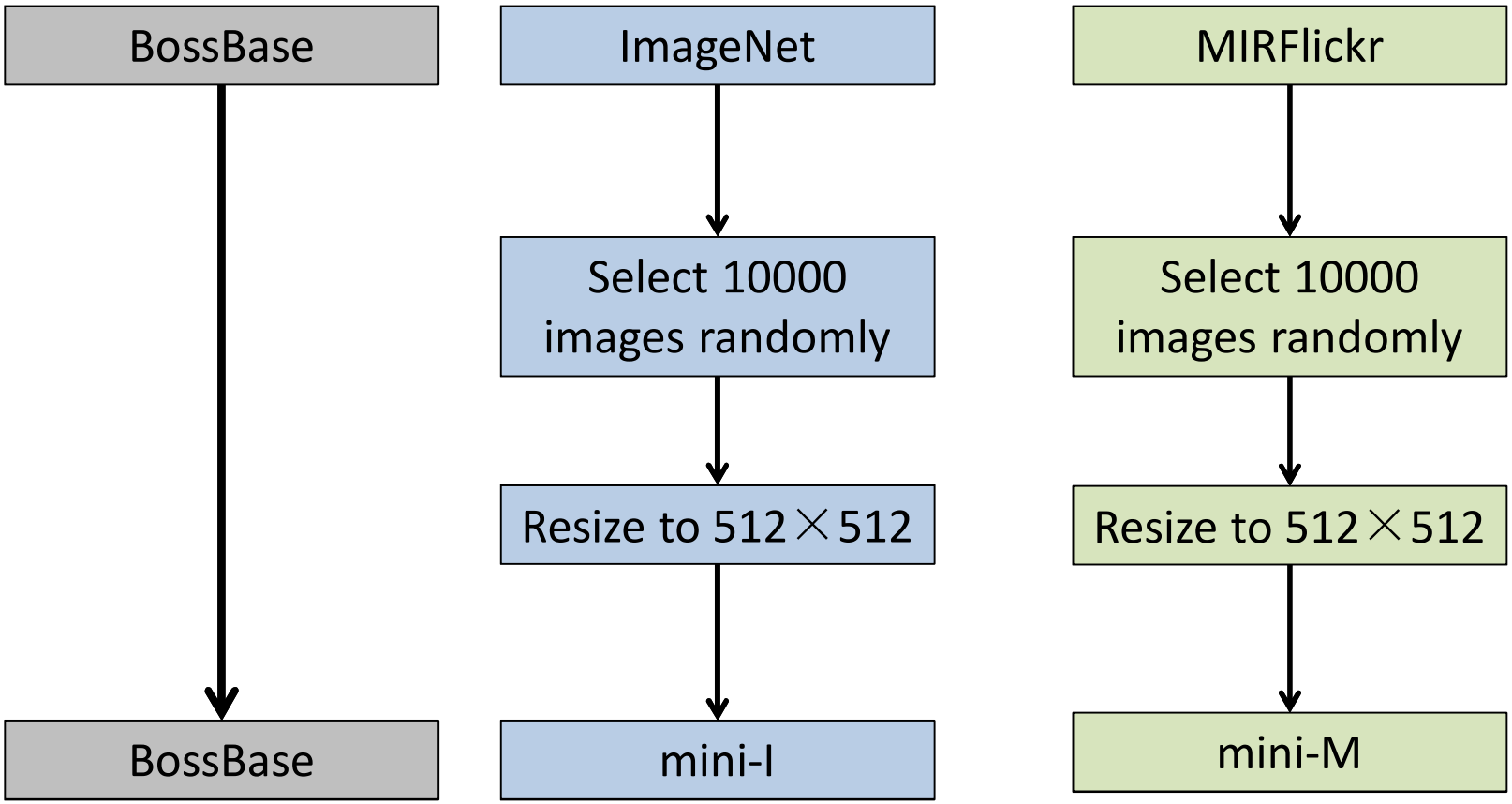
2. Research

- **1st Step**

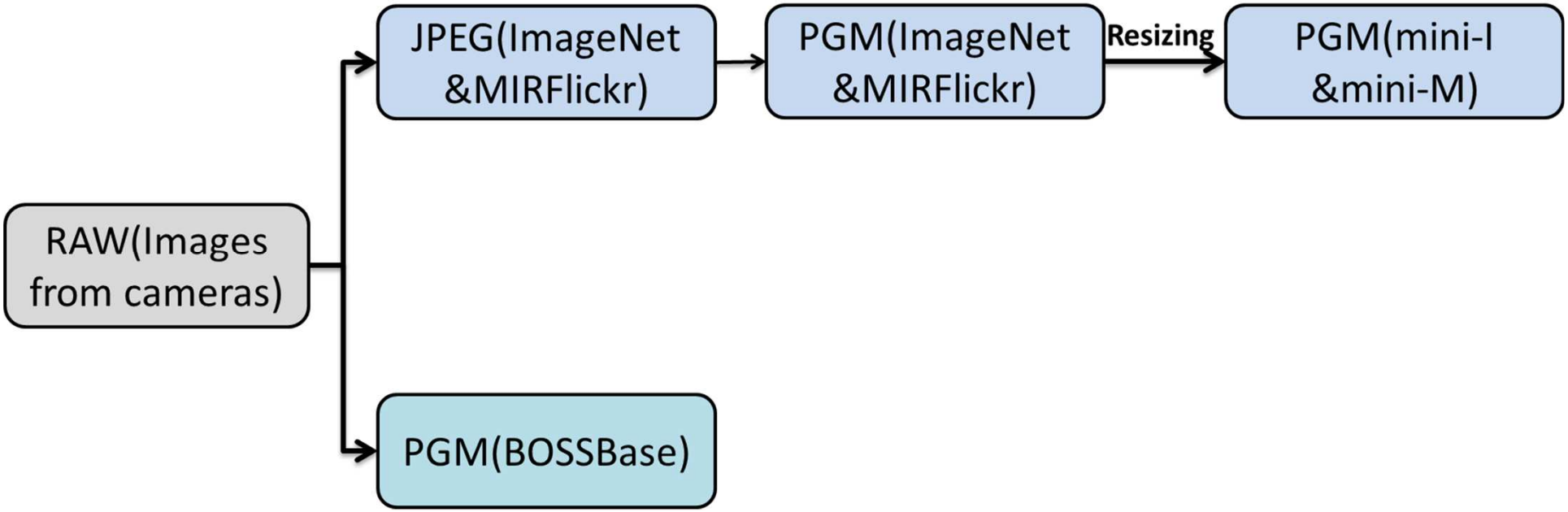
Whether there is cover-source mismatch in deep steganalysis?

BOSSBase	Common steganalysis dataset
ImageNet	Good samples from real world scenario
MIRFlickr	

2.1 Data processing



2.1 Texture complexity

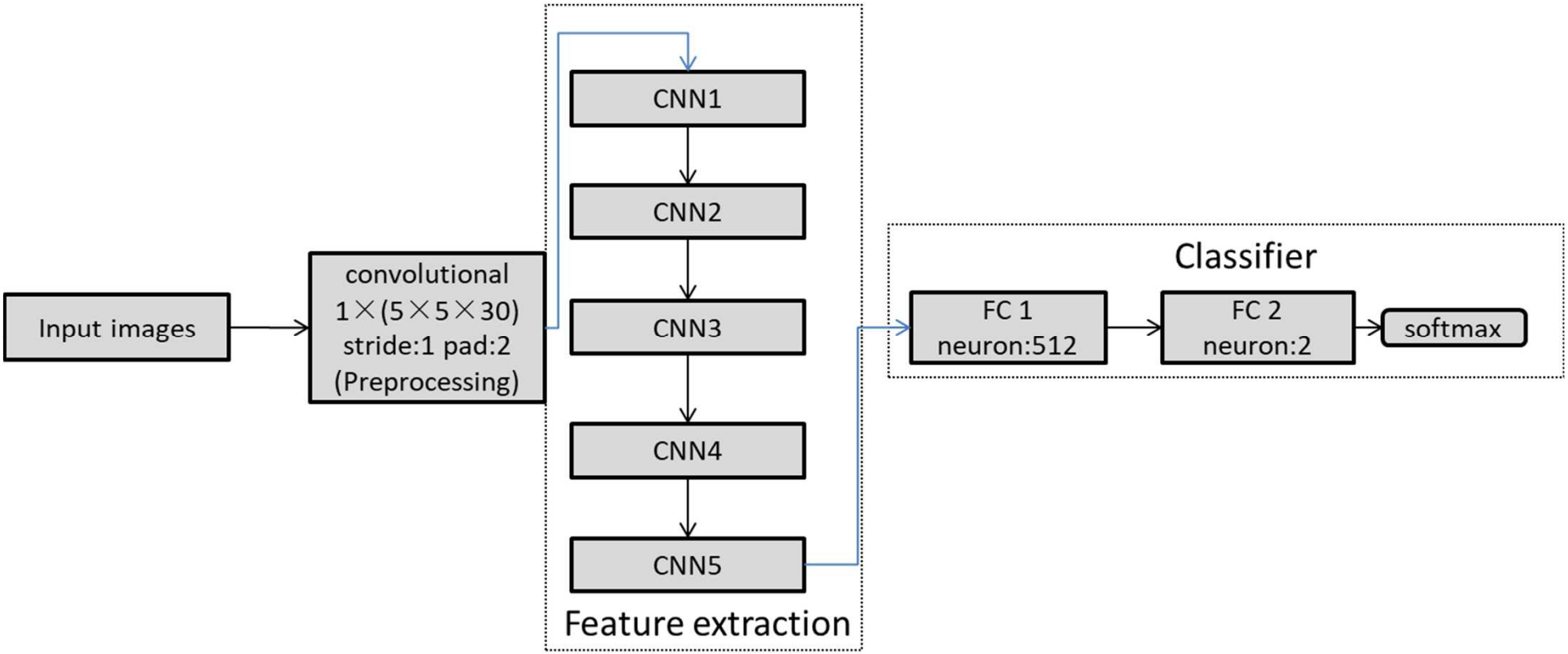


Information loss: RAW->JPEG > RAW->PGM

BOSSBase is more textured than mini-I & mini-M.

BOSSBase mini-I mini-M

2.2 Deep steganalysis model



$$\min_{f} \frac{2}{n_s} \sum_{i=1}^{n_s} (\mathrm{J}(f(\mathrm{x}_i), \mathrm{y}_i))$$

Experimental Results

train:BOSSBase			
	BOSSBase	mini-M	mini-I
suni-0.4	81.3	79.9	87.25
wow-0.4	83.323	78.15	85.475
train:mini-M			
	BOSSBase	mini-M	mini-I
suni-0.4	54.425	97.975	94.2
wow-0.4	53.825	97.875	95.675
train:mini-I			
	BOSSBase	mini-M	mini-I
suni-0.4	61.85	93.275	97.325
wow-0.4	63.175	92.45	96.475

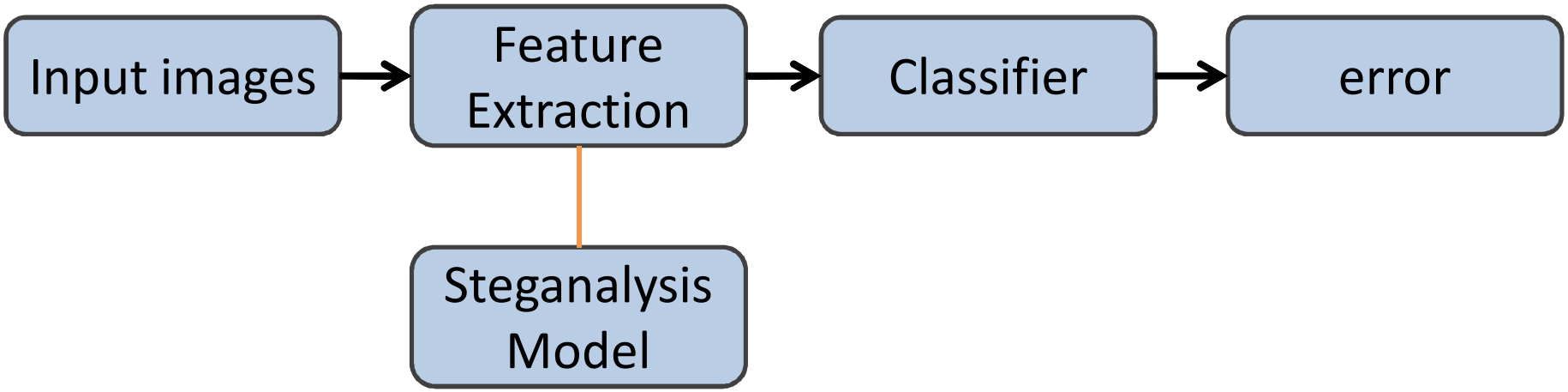
Sharp accuracy decreasing by Cover-source mismatch!

2.3 A-distance

A linear form of binary classifier error

$$\hat{d}_A = 2\left(1 - 2 \times error\right)$$

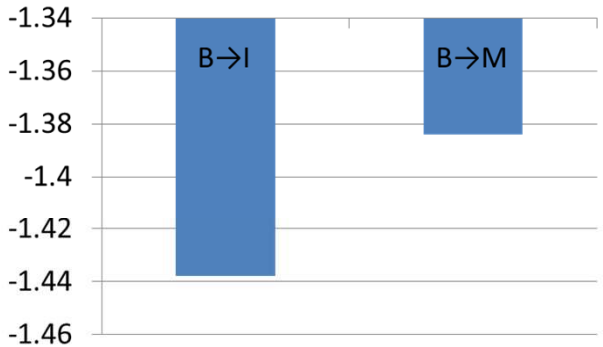
measure the discrepancy between 2 databases in the latent space



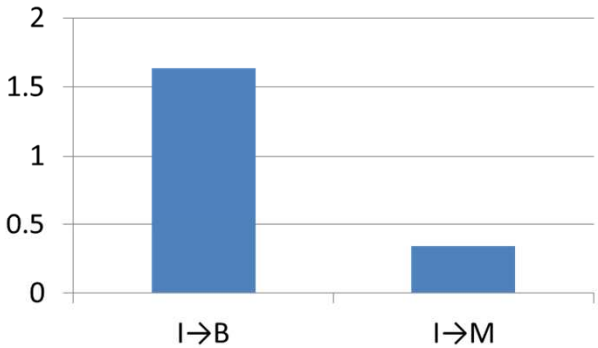
2.3 A-distance

train:BOSSBase			
	BOSSBase	mini-M	mini-I
suni-0.4	81.3	79.9	87.25
wow-0.4	83.323	78.15	85.475
train:mini-M			
	BOSSBase	mini-M	mini-I
suni-0.4	54.425	97.975	94.2
wow-0.4	53.825	97.875	95.675
train:mini-I			
	BOSSBase	mini-M	mini-I
suni-0.4	61.85	93.275	97.325
wow-0.4	63.175	92.45	96.475

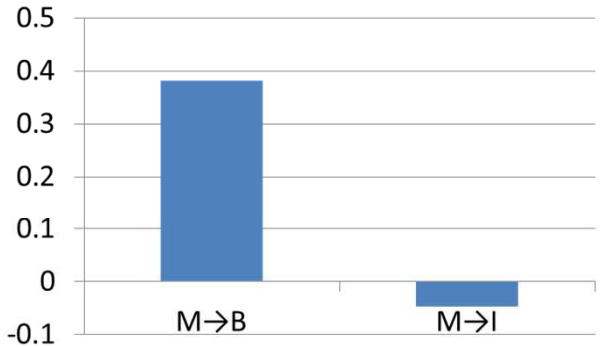
Match the experimental results well



a



b



c

Outlines

3.1 Domain adaptation

Transfer the model trained on labeled source database to unlabeled target database without sharp accuracy reduction

similar

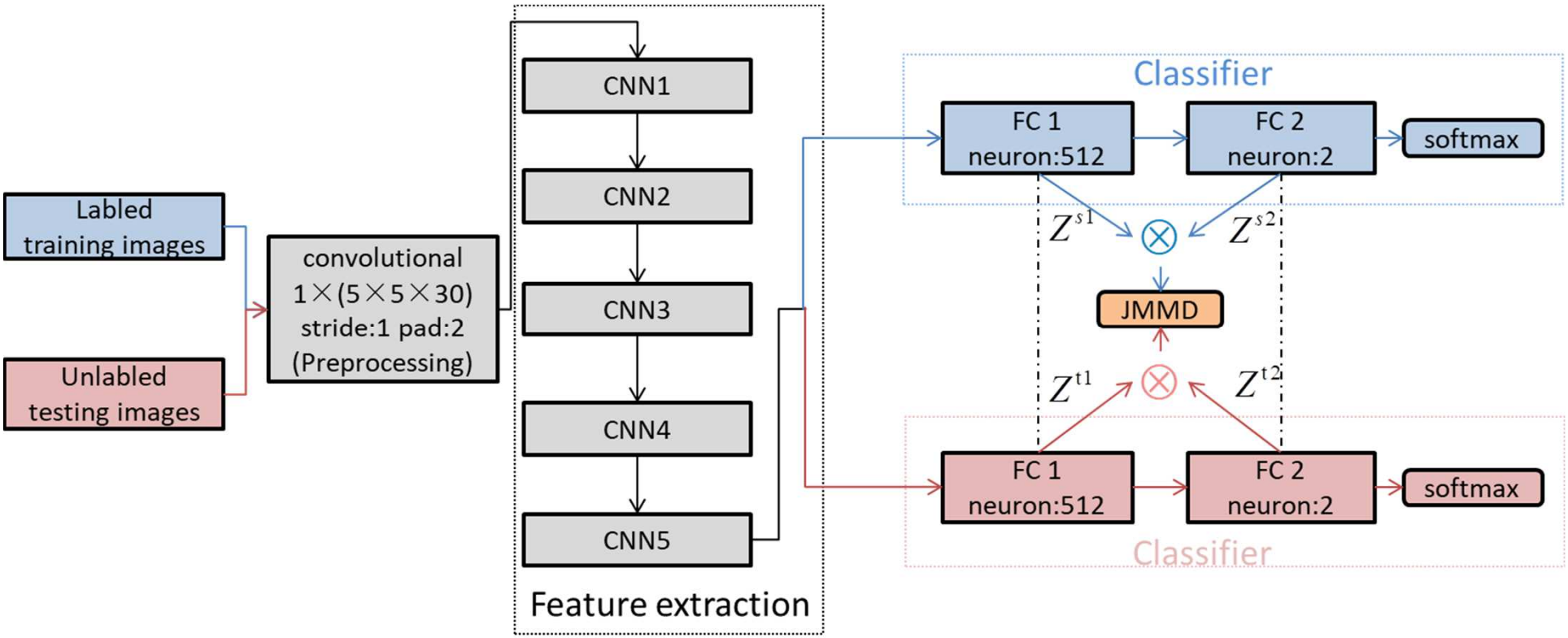
Cover-source mismatch in steganalysis

JMMD(Joint Maximum mean discrepancy):

$$D_L(P,Q) \triangleq ||L_{Z^{s,1:|L|}}(P) - L_{Z^{t,1:|L|}}(Q)||^2_{\otimes_{l=1}^{|L|} H^l}$$

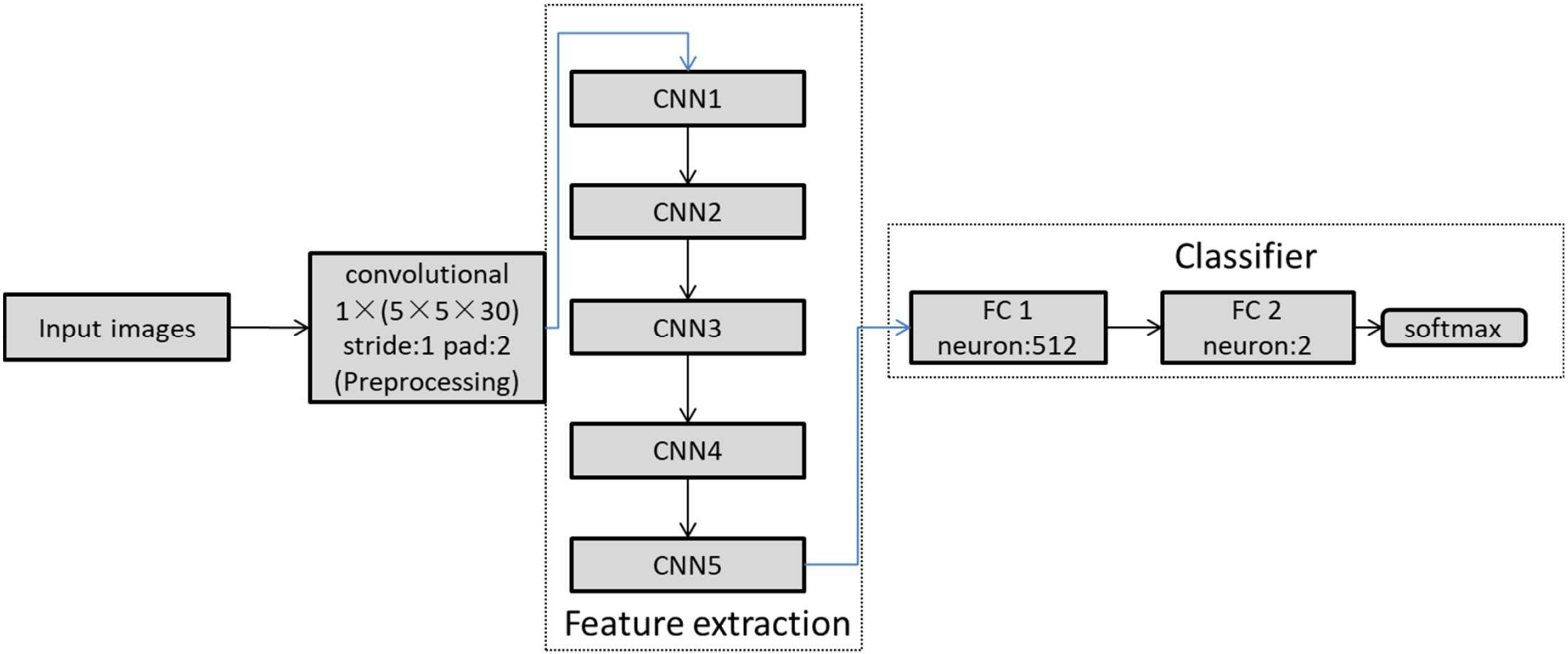
Measure and restrict the discrepancy between source and target domain in reproducing kernel Hilbert space

3.2 J-Net



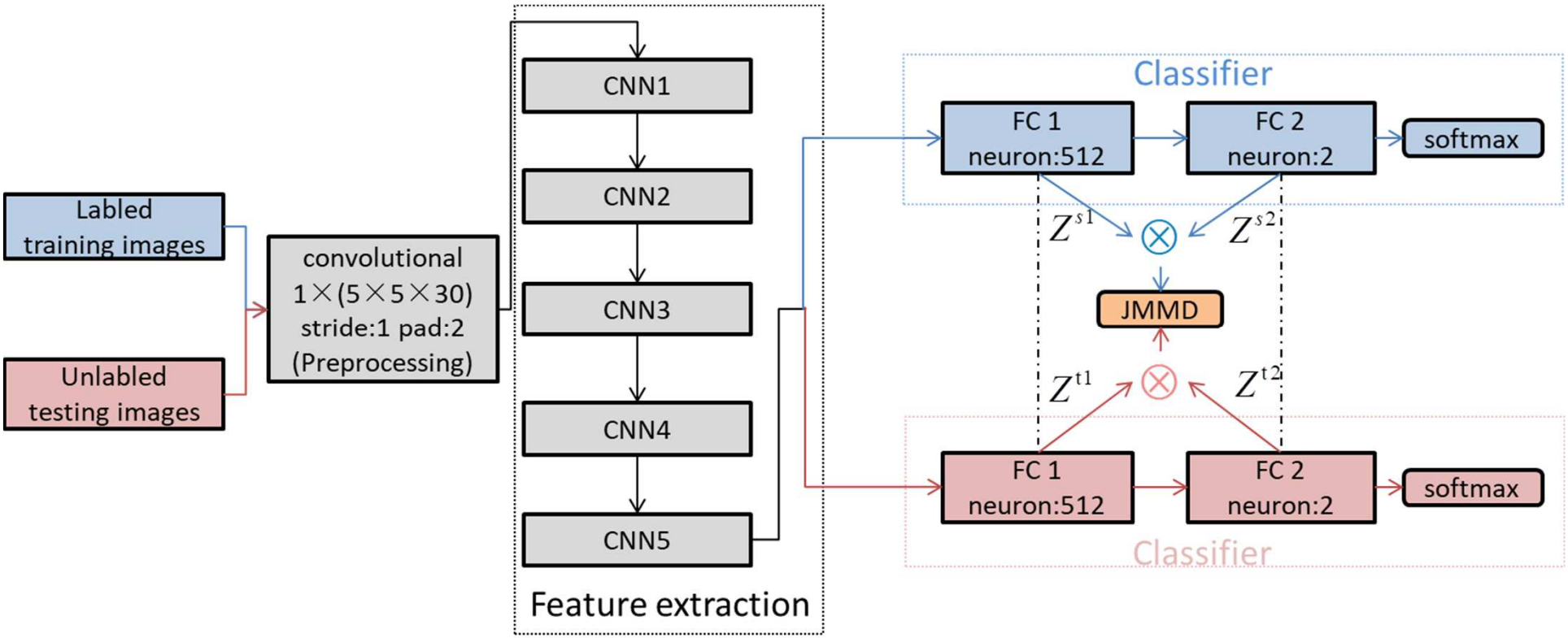
$$\min_f \frac{2}{n_s} \sum_{i=1}^{n_s} (J(f(x_i^s), y_i^s)) + \lambda \hat{D}_L(P, Q)$$

Deep steganalysis model



$$\min_f \frac{2}{n_s} \sum_{i=1}^{n_s} (J(f(x_i), y_i))$$

3.2 J-Net



$$\min_f \frac{2}{n_s} \sum_{i=1}^{n_s} (J(f(x_i^s), y_i^s)) + \lambda \hat{D}_L(P, Q)$$

3.3 Results

The accuracy promotion of J-Net(%)

train:mini-I test:BOSSBase			
	pre-train	J-Net	promotion
suni-0.4	61.85	68.95	7.1
wow-0.4	63.175	71.2	8.025
train:mini-I test:BOSSBase			
	pre-train	J-Net	promotion
suni-0.4	54.425	63.875	9.45
wow-0.4	53.825	63.725	9.9

7%-10%!

THANK YOU!

Presented by Xunpeng Zhang