#### Stochastic Label Refinery: Toward Better Target Label Distribution

Xi Fang, Jiancheng Yang, Bingbing Ni

Shanghai Jiao Tong University





## **Target Distribution**

Which is better target label distibution?

	Categories		Sky	Cloud	Sea	Plant	Others
2000	Hord Labol	one-hot	1	0	0	0	0
		multi-hot	1	1	0	1	0
	Coft Lobal	label smoothing	0.9	0.025	0.025	0.025	0.025
et sa carse and some	Son Laber	ideal distribution	0.45	0.1	0.05	0.35	0.05

Soft is better than Hard:

- · hard: over-confidence (one-hot)
- · soft: better performance (label smoothing) , more information (knowledge distillation)
- $\cdot$  soft: robust to noise-label or long-tail

# **Stochastic Label Refinery**

Each round of SLR:



Use out-of-fold (oof) pseudo-label to refine the target label distribution by simply weighted average

out-of-fold (oof): https://machinelearningmastery.com/out-of-fold-predictions-in-machine-learning/

## **Stochastic Label Refinery**



Using better target label distribution we can get better models by multiple rounds SLR (Self-distillation)

- $\cdot$  K-fold teachers and K-fold students using the same network architecture
- · Each round of K-fold is divided differently

#### Experiments

In CIFAR-10 :

Using blending method in self knowledge distillation to get better target label ditribution (SLR-a/e): We denote that better target label distribution leads to better socre.

Method	Top-1 accuracy
VGG16 [37]	92.64%
ResNet101 [38]	93.75%
DenseNet121 [39]	95.04%
PreResNet56 [40]	95.51%
SE-ResNet56 [6]	95.87%
SE-ResNet56 [6] + AA [3]	96.16%
SE-ResNet56 [6] + AA [3] + Label Smoothing [12]	96.16%
SE-ResNet56 [6] + AA [3] + SLR	96.41%
SE-ResNet56 [6] + AA [3] + SLR-a	96.44%
SE-ResNet56 [6] + AA [3] + SLR-e	96.42%
SE-ResNet56 [6] + AA [3] + SLR-ae	96.53%

### Experiments

#### Using SLR we get SOTA in DeepDR Diabetic Retinopathy Dataset

Method	Quadratic Weighted Kappa	
Baseline (w/o tricks)	$0.8036 \pm 0.0214$	
Baseline (w/ tricks)	$0.8247 \pm 0.0125$	
SWA [34]	$0.8119 \pm 0.0234$	
OHEM [20]	$0.8061 \pm 0.0174$	
Knowledge Distillation [27]	$0.8128 \pm 0.0100$	
Label Refinery [16]	$0.7527 \pm 0.0152$	
Stochastic Label Refinery	$0.8348 {\pm} 0.0053$	

Method	Public Test	Private Test 0.9215	
Ours	0.9303		
Team1	0.9262	0.9211	
Team2	0.9232	0.9097	
Team3	0.9202	0.8946	
Team4	0.9088	0.8890	

improve from high baseline with multiple training tricks







better than vanilla label refinery

#### How does SLR works?

• As a regularization strategy :

just like label smoothing (avoid over-confidence)

• As a label correction method :







round 7



round 8



SLR (8 rounds)