Paper ID: 1948

Towards Robust Learning with Different Label Noise Distributions

Diego Ortego, Eric Arazo, Paul Albert, Noel E. O'Connor and Kevin McGuinness

Code on github
https://git.io/JJ0PV

Exploiting large data collections

- DNN training with more data → Better results
- Labeling large data collections is expensive
- Possible solution: collect web images and infer the labels from the metadata

✅ Low human supervision

❌ Uncontrolled label noise

Label noise

Studied in a controlled manner by flipping labels to incorrect classes (synthetic in-distribution (ID) noise)

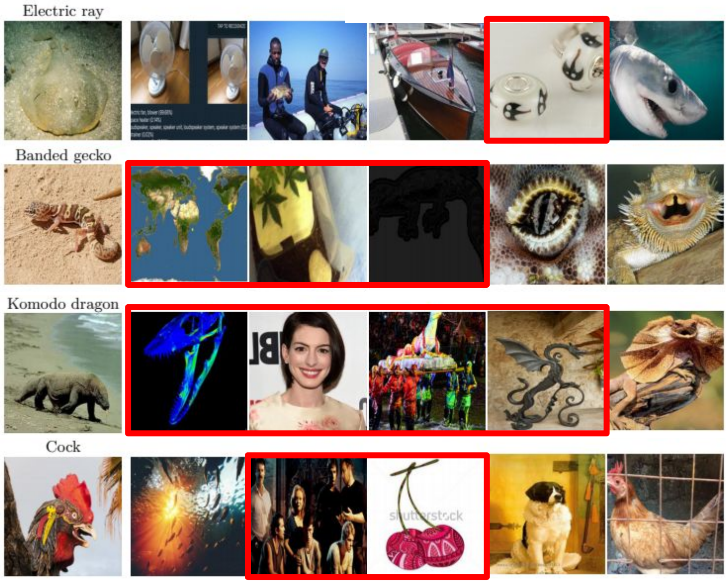
Methods designed to perform label correction

[1] Li et al. WebVision Database: Visual Learning and Understanding from Web Data. arXiv 2017

Label noise

Studied in a controlled manner by flipping labels to incorrect classes (synthetic in-distribution (ID) noise)

Methods designed to perform label correction



But, in real-world noise…

(WebVision [1] examples)

Out-of-distribution (OOD) noise **highly present**

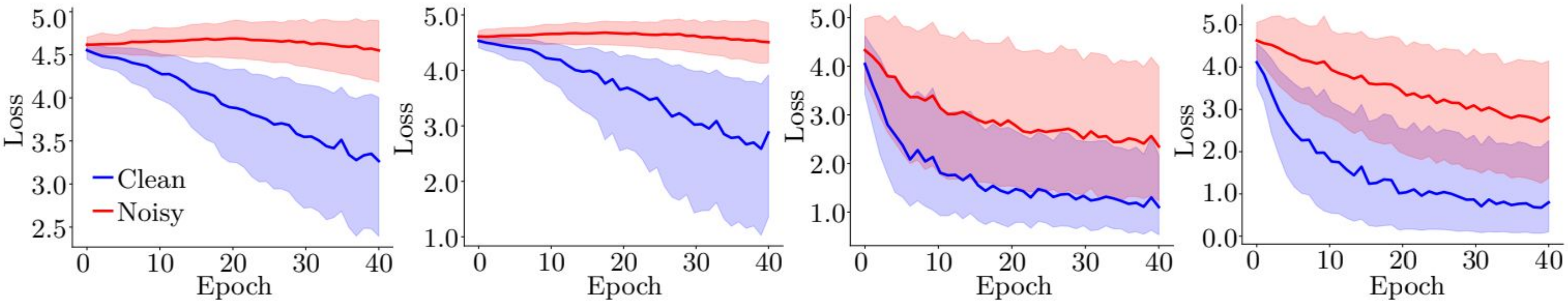
Important to consider different noise types

[1] Li et al. WebVision Database: Visual Learning and Understanding from Web Data. arXiv 2017

Impact of label noise distribution

Small loss trick (low loss=clean) often used is **not straightforward** to apply for all **noise distributions**

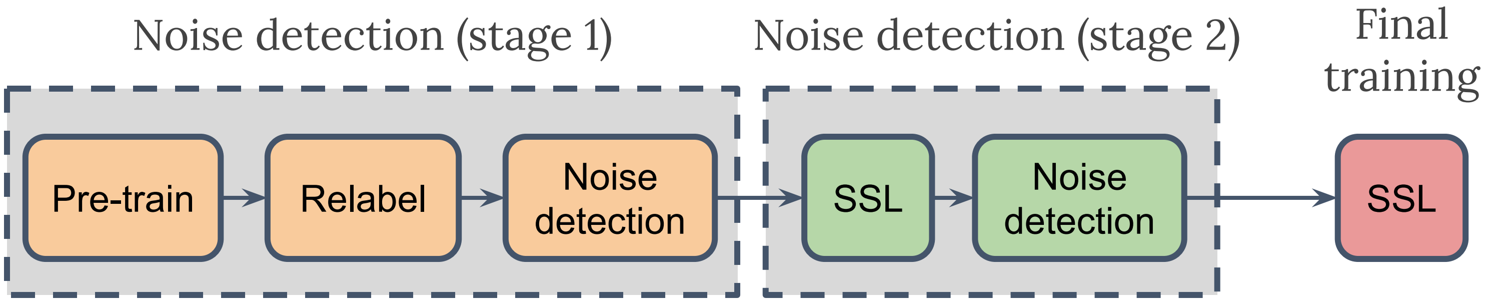
Uniform ID | Uniform OOD | Non-uniform ID | Non-uniform OOD



Proposed method

Label noise **Distribution Robust Pseudo-Labeling** (DRPL):

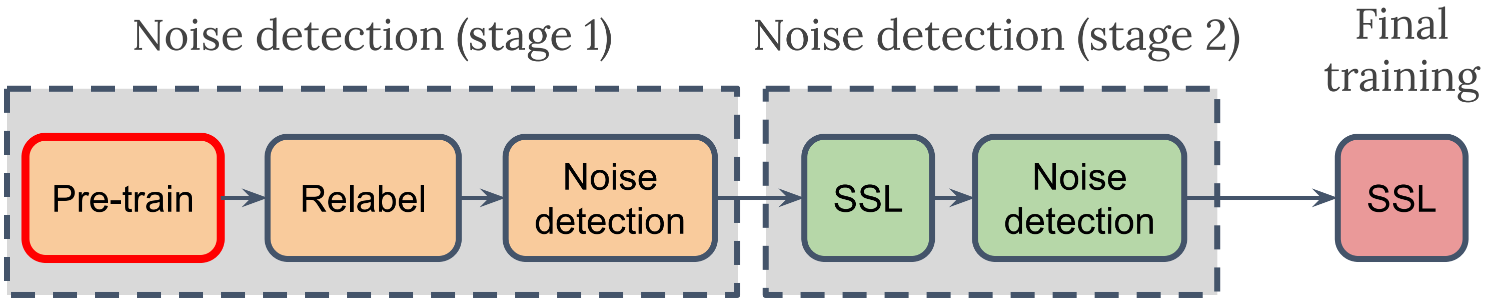
1) Label noise detection. 2) Semi-supervised learning (SSL)



Proposed method

Label noise **Distribution Robust Pseudo-Labeling** (DRPL):

1) Label noise detection. 2) Semi-supervised learning (SSL)

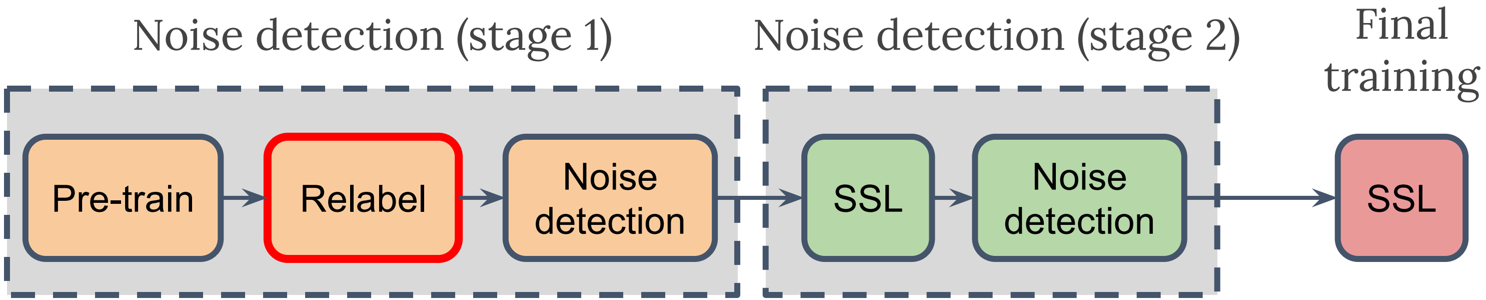


Pre-train with high learning rate to learn clean data pattern without memorizing label noise

Proposed method

Label noise **Distribution Robust Pseudo-Labeling** (DRPL):

1) Label noise detection. 2) Semi-supervised learning (SSL)



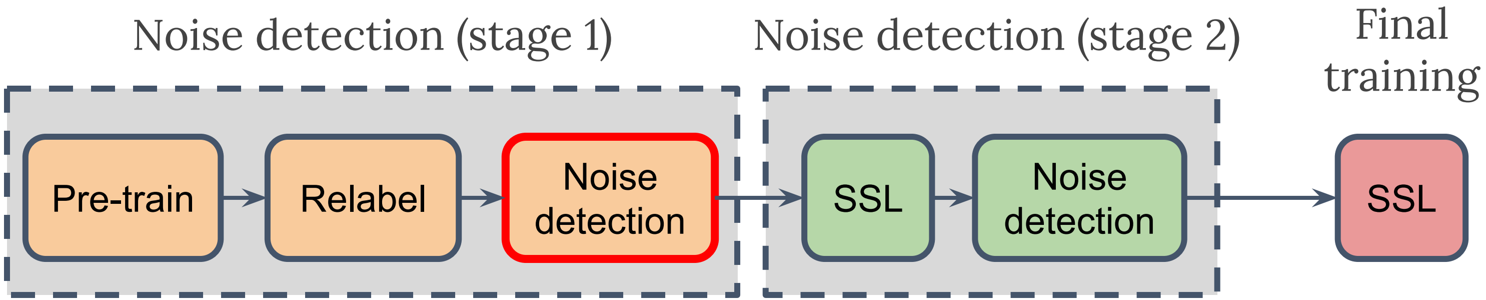
Relabel all samples with network predictions [2]. This stage is key to reveal a discriminative measure for noise detection.

[2] Tanaka et al. [Joint Optimization Framework for Learning with Noisy Labels](). CVPR 2018

Proposed method

Label noise **Distribution Robust Pseudo-Labeling** (DRPL):

1) Label noise detection. 2) Semi-supervised learning (SSL)



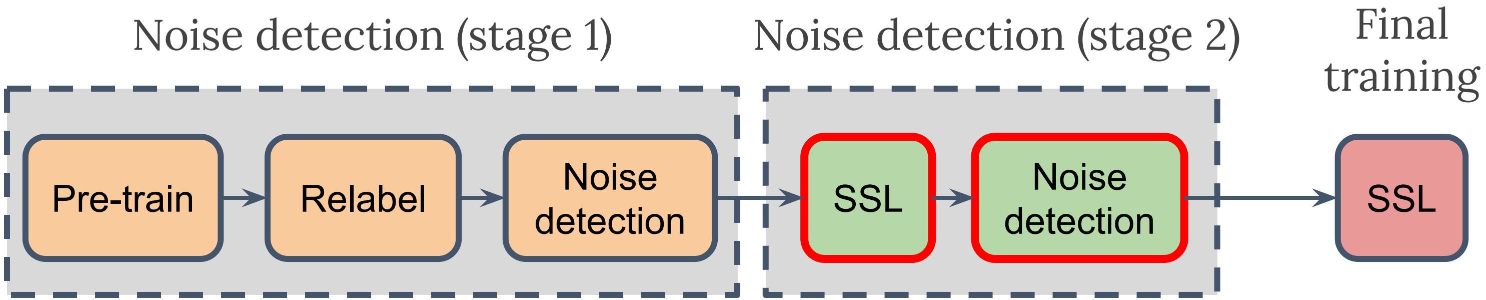
Noise detection based on fitting a Beta mixture model [3] to the discriminative measure and thresholding the posterior distribution

[3] Arazo et al. Unsupervised Label Noise Modeling and Loss Correction. ICML 2019

Proposed method

Label noise **Distribution Robust Pseudo-Labeling** (DRPL):

1) Label noise detection. 2) Semi-supervised learning (SSL)



SSL using [4] where clean=labeled and noisy=unlabeled

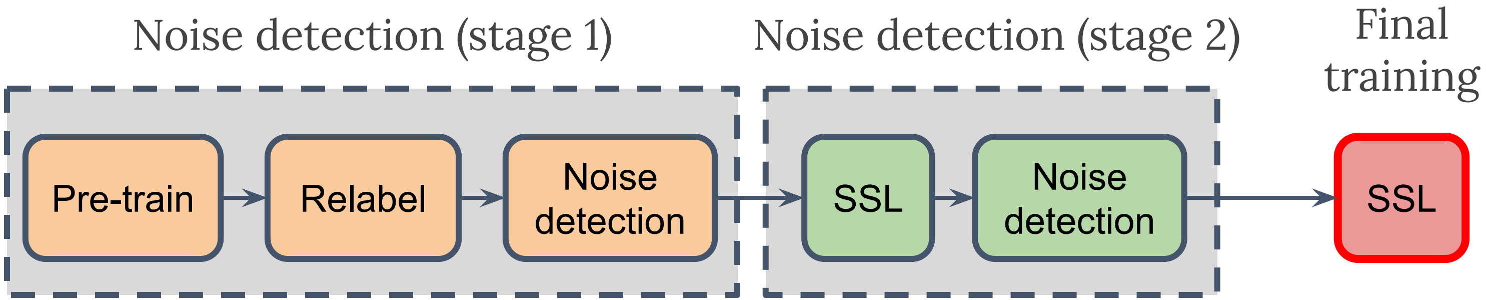
Repeat noise detection to refine the clean and noisy sets

[4] Arazo et al. Pseudo-Labeling and Confirmation Bias in Deep Semi-Supervised Learning. IJCNN 2020

Proposed method

Label noise **Distribution Robust Pseudo-Labeling** (DRPL):

1) Label noise detection. 2) Semi-supervised learning (SSL)



Final SSL [4] training with the refined clean/noisy sets

[4] Arazo et al. [Pseudo-Labeling and Confirmation Bias in Deep Semi-Supervised Learning](). IJCNN 2020

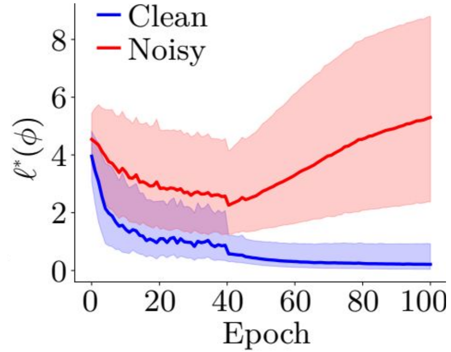
Discriminative measure during relabeling

Cross-entropy (CE) between current predictions and old labels

=

Disagreement between the new and old noise patterns

Clean (Low CE)
Keep the same labels in both patterns
Noisy (High CE)
Exhibit different labels in both patterns



Evaluation (ImageNet32/64, CIFAR-10/100, mini-WebVision)

Uniform (U) and non-uniform (NU) for ID and OOD noise

	ImageNet32								ImageNet64							
	NU-ID		U-ID		NU-OOD		U-OOD		NU-ID		U-ID		NU-OOD		U-OOD	
	30%	50%	40%	80%	30%	50%	40%	80%	30%	50%	40%	80%	30%	50%	40%	80%
FW [24]	54.22	43.38	52.06	31.20	62.14	55.06	56.32	40.08	60.10	46.06	57.42	37.84	69.86	63.38	63.08	47.68
R [14]	67.24	63.62	62.98	41.52	66.36	62.80	64.04	45.00	74.28	69.20	70.98	48.44	74.22	70.74	72.78	54.00
M [40]	67.14	51.96	61.98	38.92	66.14	60.62	64.66	47.40	74.02	58.14	69.90	49.22	74.78	69.40	73.94	59.54
DB [16]	62.88	52.20	67.62	45.34	64.86	60.58	65.96	39.30	71.30	60.98	74.56	56.44	77.94	70.38	74.08	50.98
DRPL	**73.46**	**68.18**	**73.48**	**61.78**	**71.38**	**67.32**	**71.36**	**54.10**	**81.90**	**77.66**	**81.50**	**73.08**	**80.44**	**76.38**	**79.76**	**64.34**

mini-WebVision

CE	FW [24]	R [14]	M [40]	GCE [35]	DB [16]	DMI [17]	P [15]	DM [38]	DRPL
73.88	74.68	76.52	80.76	74.28	79.68	73.96	79.96	78.16	**82.08**

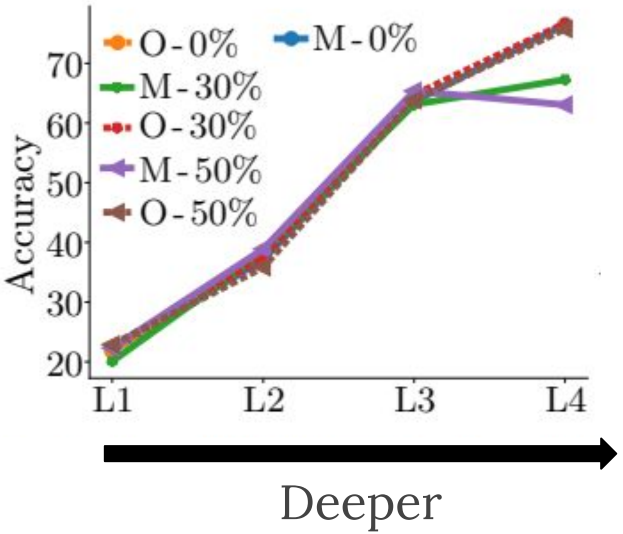
✅State-of-the-art results
✅Consistent across noise levels, distributions, and image resolutions

Real-world (RW) vs synthetic (SY)

Take home message:
Most SY noises are different from RW noise
RW noise contains OOD samples
OOD samples should be treated differently (mixup success)

Representation learning

ImageNet64



- M: Mixup, O: Ours (DRPL)
- Linear model evaluation for features at different depths
- Degradation concentrates at the end

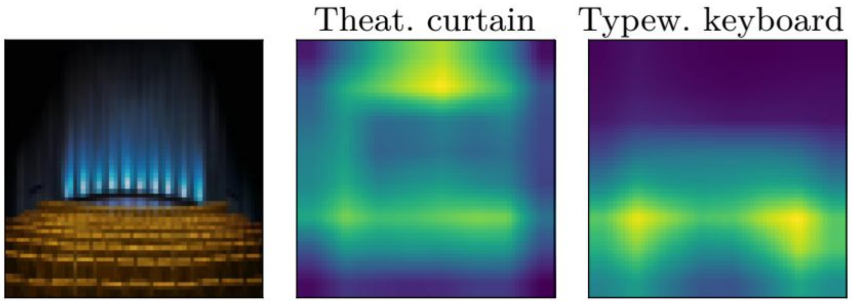
Take home message: Even if label noise is memorized, discriminative low-level and mid-level features emerge

Where DNNs look for memorizing?

Activation maps

Noisy label: Keyboard

True label: Theater curtain



Activation map curtain class (not predicted)

Activation map keyboard class (predicted)

Take home message: DNNs skip relevant areas for the true class, while focusing on areas that help explaining the noisy label

Conclusions

- DRPL: robust image classification models in the presence of label noise
- Robustness comes from an effective label noise detection for different noise distributions
- We analyze different label noise distributions from multiple perspectives (similarities with RW noise, representation learning, attention maps...) leading to important conclusions that help in better understanding label noise effect

Thank you!

Q/A?