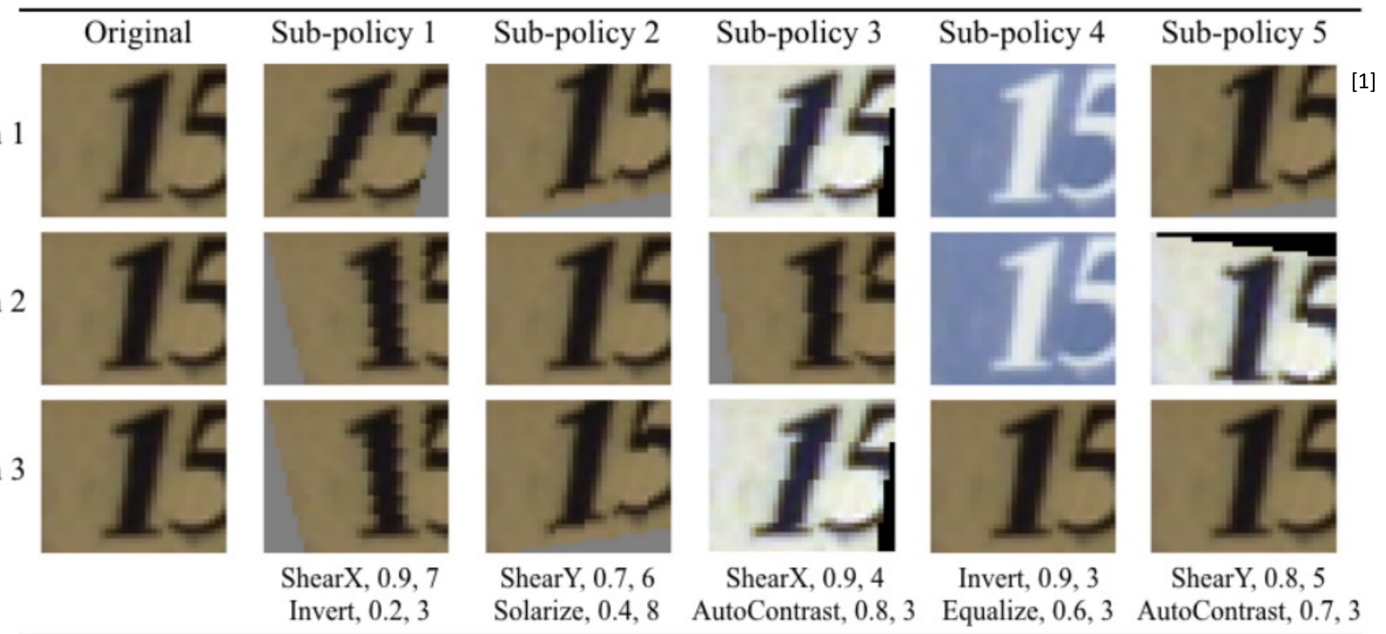
Adversarial Auto-Augment with Label Preservation: A Representation Learning Principle Guided Approach

Kaiwen Yang, Yanchao Suns, Jiahao Su, Fengxiang He, Xinmei Tian, Furong Huang, Tianyi Zhou, Dacheng Tao

Problem of current data augmentations

- They are based on pre-defined operations and are **not** fully automated.



- They rely on domain knowledge to preserve label and are thus **restricted to few domains**.

Medical Images

Their RandAugment

Preliminary: representation learning with data augmentation

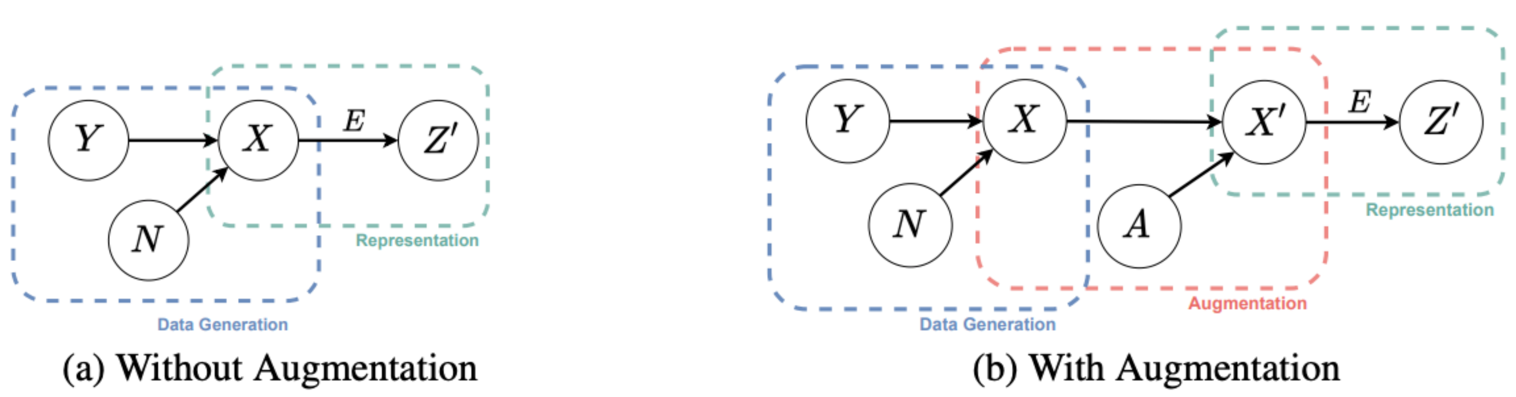


Figure 1: **Probabilistic graphical models** of representation learning.

X: data observation

Y: label

N: nuisance part in data, which is independent to label

Z': low-dimensional representation of data (mapped by an encoder E)

A: augmentation selection

X': augmented data

[1] AutoAugment: Learning Augmentation Policies from Data. CVPR 2019.

[2] MedMNIST v2 -- A large-scale lightweight benchmark for 2D and 3D biomedical image classification, Arxiv 2021

What is a good representation?

Definition 4.0.1 (ϵ-Minimal Sufficient Representation (ϵ-Optimal Representation)). *For a Markov chain $\mathbf{Y} \to \mathbf{X} \to \mathbf{Z}$, we say thata representation $\mathbf{Z}$ of $\mathbf{X}$ is sufficient for $\mathbf{Y}$ if $I(\mathbf{Z} \wedge \mathbf{Y}) = I(\mathbf{X} \wedge \mathbf{Y})$, and $\mathbf{Z}$ is ϵ-minimal sufficient for $\mathbf{Y}$ if $\mathbf{Z}$ is sufficient and $I(\mathbf{Z} \wedge \mathbf{X}) \le I(\tilde{\mathbf{Z}} \wedge \mathbf{X}) + \epsilon$ for all $\tilde{\mathbf{Z}}$ satisfying $I(\tilde{\mathbf{Z}} \wedge \mathbf{Y}) = I(\mathbf{X} \wedge \mathbf{Y})$.*

Sufficiency: should contain **all** the information about label Y.

Minimality: should contain as **little** information as possible about data X.

Proper data augmentation leads to optimal representation

Theorem 4.2. *Consider label variable $\mathbf{Y}$, observation variable $\mathbf{X}$ and nuisance variable $\mathbf{N}$ satisfying Assumption 4.1. Let $\mathbf{A}$ be the augmentation variable, $\mathbf{X}'$ be the augmented data, and $\mathbf{Z}^*$ be the solution to*

$$\begin{aligned} &\operatorname{argmax}_{\mathbf{Z}'} \quad I(\mathbf{Z}' \wedge \mathbf{X}') \text{ or } I(\mathbf{Z}' \wedge \mathbf{Y}) \\ &\text{subject to} \quad I(\mathbf{Z}' \wedge \mathbf{A}) = 0. \end{aligned} \tag{1}$$

Then, $\mathbf{Z}^*$ is a ϵ-minimal sufficient representation of $\mathbf{X}$ for label $\mathbf{Y}$ if the following conditions hold:
Condition (a): *$I(\mathbf{X}' \wedge \mathbf{Y}) = I(\mathbf{X} \wedge \mathbf{Y})$ ($\mathbf{X}'$ is an in-class augmentation) and*
Condition (b): *$I(\mathbf{X}' \wedge \mathbf{N}) \le \epsilon$ ($\mathbf{X}'$ does not remain much information about $\mathbf{N}$).*

Label-preservation: keep all the label-relevant information in augmentation

Adversary: maximally perturbs the label-irrelevant information

Label-Preserving Adversarial Auto-Augment (LPA3)

Initial optimization problem:

$$\min_{\mathbf{X}'} I(\mathbf{X}' \wedge \mathbf{X}) \quad \text{s.t. } I(\mathbf{X}' \wedge \mathbf{Y}) = I(\mathbf{X} \wedge \mathbf{Y}).$$

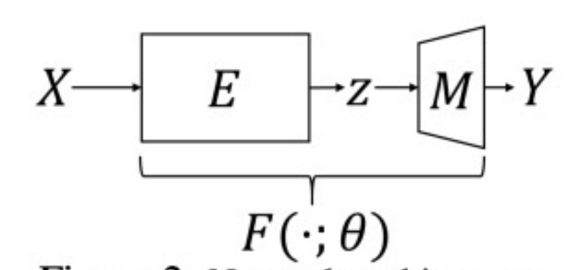


Figure 2: Network architecture.

Implementation of mutual information:

Constraint term: $\log F(x';\theta)[y] = \log F(x;\theta)[y]$ (Neural network classification result)

Objective term: $\mathrm{LPIPS}(x,x') \triangleq \|\phi(x) - \phi(x')\|_2$. (Neural network midden-layer features)

The final optimization problem:

$$\min_{x'} -\|\phi(x) - \phi(x')\|_2 + \lambda \max(0, \log F(x;\theta)[y] - \log F(x';\theta)[y] - \sigma)$$

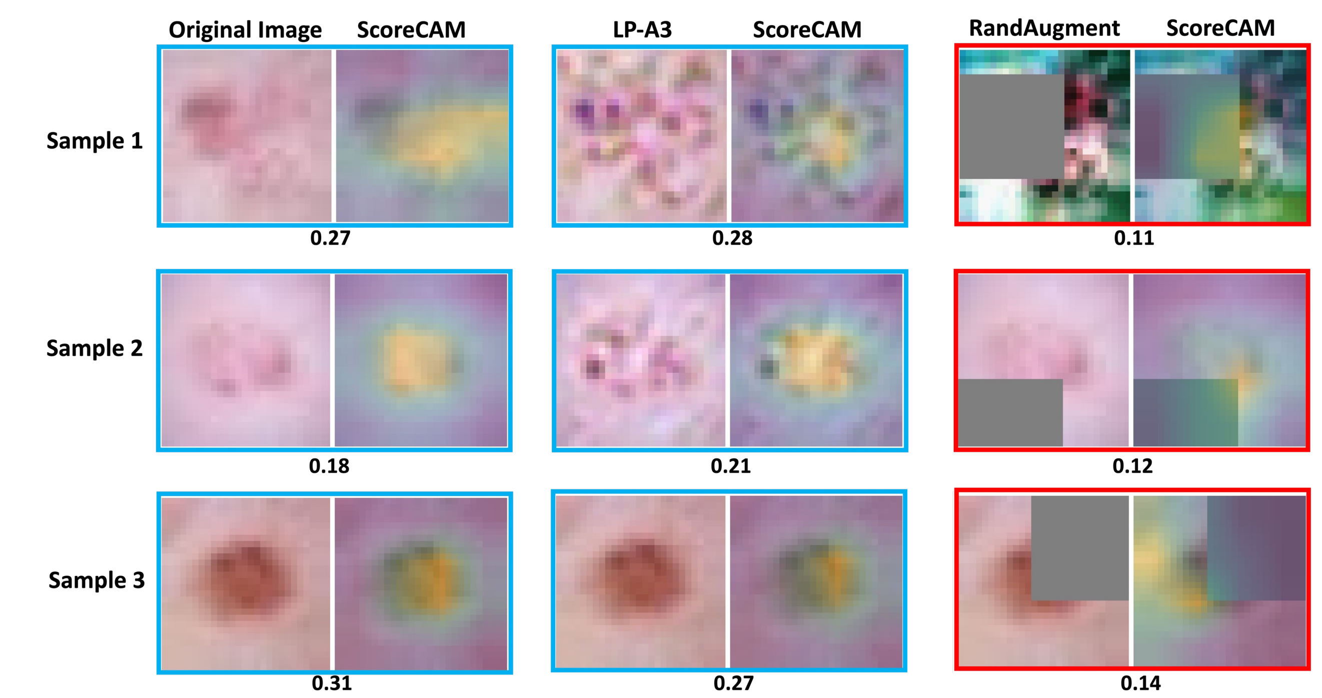
Algorithm 1 Plug LP-A3 into any representation learning procedure

Input: Loss for the targeted task $L : \mathcal{X} \times \mathcal{Y} \times \mathcal{W} \to \mathcal{R}_+$; training data $(\mathcal{X}, \mathcal{Y})$; neural network $F(\cdot;\theta)$; class preserving margin ϵ; data selection ratio τ; learning rate η;
Output: Model parameter θ trained with LP-A3

1: **while** *not converged* **do**
2: Sample batch $\mathcal{B} = \{(x_1, y_1), ..., (x_b, y_b)\} \sim (\mathcal{X}, \mathcal{Y})$;
3: Data selection: $\mathcal{S} \leftarrow \tau\%$ data with the lowest TCS in $\mathcal{B}$;
4: LP-A3: Freeze θ and solve Equation (5) using Algorithm 2 for every sample in $\mathcal{S}$, resulting in an augmented set $\mathcal{A} = \{(x'_1, y_1), ..., (x'_m, y_m)\}$ of size $m = |\mathcal{S}|$;
5: Learning with LP-A3 augmented data and original data: $\theta \leftarrow \theta - \eta[\nabla_\theta L(\mathcal{B};\theta) + \nabla_\theta L(\mathcal{A};\theta)]$;
6: **end while**

Experimental Results

Visualization on MedMNIST



Semi-supervised learning

Dataset	CIFAR10			CIFAR100			STL-10
# Label	40	250	4000	400	2500	10000	1000
InfoMin (RGB) [40]	-	-	-	-	-	-	86.0
InfoMin (YDbDr) [40]	-	-	-	-	-	-	87.0
FixMatch [36]§	89.51±3.14	93.81±0.29	94.66±0.13	49.30±2.45	67.21±0.94	74.31±0.35	91.59±0.16
FixMatch [36] + LP-A3	**92.39±1.21**	**94.03±0.31**	**95.11±0.17**	**56.16±1.82**	**72.23±0.57**	**77.11±0.16**	**92.63±0.14**

Noisy-label learning

Dataset	CIFAR10			CIFAR100		
Noise Ratio	50%	80%	90%	50%	80%	90%
Mixup [56]	87.1	71.6	52.2	57.3	30.8	14.6
P-correction [54]	88.7	76.5	58.2	56.4	20.7	8.8
M-correlation [3]	88.8	76.1	58.3	58.0	40.1	14.3
DivideMix [26]	94.4	92.9	75.4	74.2	59.6	31.0
DivideMix+LP-A3	94.89±0.05	**93.70±0.19**	79.35±1.33	74.12±0.23	61.00±0.34	32.55±0.25
PES§ [5]	94.89±0.12	92.15±0.23	84.98±0.36	74.19±0.23	61.47±0.38	21.15±3.15
PES+LP-A3	**95.10±0.14**	93.26±0.21	**87.71±0.36**	**74.57±0.25**	**62.98±0.49**	**40.61±1.10**

Medical image classification

Method	PathMNIST	DermaMNIST	TissueMNIST	BloodMNIST
ResNet-18	94.34±0.18	76.14±0.09	68.28±0.17	96.81±0.19
ResNet-18+RandAugment	93.52±0.09	73.71±0.33	62.03±0.14	95.00±0.21
ResNet-18+LP-A3	**94.42±0.24**	**76.22±0.27**	**68.63±0.14**	**96.97±0.06**
ResNet-50	94.47±0.38	75.24±0.27	69.69±0.23	96.91±0.06
ResNet-50+RandAugment	94.02±0.37	71.65±0.30	65.13±0.33	95.14±0.06
ResNet-50+LP-A3	**94.57±0.07**	**75.71±0.22**	**69.89±0.08**	**97.01±0.32**
	OctMNIST	OrganAMNIST	OrganCMNIST	OrganSMNIST
ResNet-18	78.67±0.26	94.21±0.09	91.81±0.12	81.57±0.07
ResNet-18+RandAugment	76.00±0.24	94.18±0.20	91.38±0.14	80.52±0.32
ResNet-18+LP-A3	**80.27±0.54**	**94.73±0.21**	**92.41±0.22**	**82.28±0.38**
ResNet-50	78.37±0.52	94.31±0.14	91.80±0.14	81.11±0.21
ResNet-50+RandAugment	76.63±0.58	94.59±0.17	91.10±0.12	80.47±0.37
ResNet-50+LP-A3	**79.40±0.36**	**94.95±0.19**	**92.16±0.23**	**82.15±0.08**