Augment with Care: Contrastive Learning for Combinatorial Problems

Haonan Duan, Pashootan Vaezipoor*, Max B. Paulus, Yangjun Ruan, Chris J. Maddison*

VECTOR INSTITUTE

UNIVERSITY OF TORONTO

ETH zürich

Supervised learning for combinatorial problems

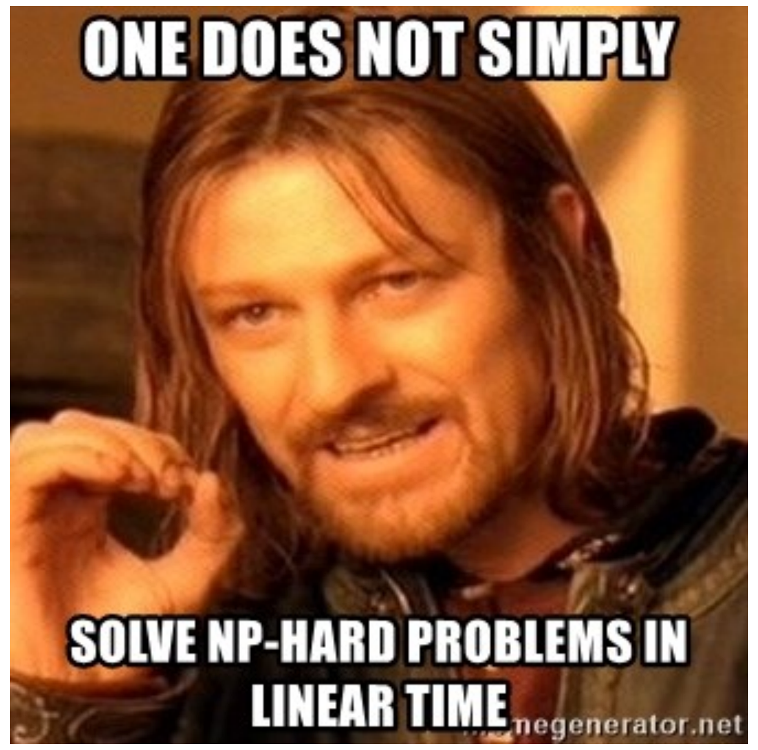
	Input	Label	ML model (e.g.,)
Boolean satisfiability (SAT)	$(x_1 \lor x_2 \lor \neg x_3)$ $\land (\neg x_1 \lor x_2 \lor x_3)$	SAT/UNSAT	NeuroSAT (Selsam et al, 2019)
Mixed Integer Programming	max $x_1 - 3x_2$ s.t. $x_1 > x_2$ $x_1, x_2 \in \mathbb{Z}$	Variable assignments	Neural Diving (Nair et al, 2020)
Travelling salesman		Shortest routes	Attention Model (Kool et al, 2018)

Supervised learning for combinatorial problems

	Input	Label	ML model (e.g.,)
Boolean satisfiability (SAT)	$(x_1 \vee x_2 \vee \neg x_3)$ $\wedge (\neg x_1 \vee x_2 \vee x_3)$	SAT/UNSAT	NeuroSAT (Selsam et al, 2019)
Mixed Integer Programming	max $x_1 - 3x_2$ s.t. $x_1 > x_2$ $x_1, x_2 \in \mathbb{Z}$	Variable assignments	Neural Diving (Nair et al, 2020)
Travelling salesman		Shortest routes	Attention Model (Kool et al, 2018)

Limitation:

- Combinatorial optimization is NP-hard
- Worst-case exponential complexity
- Labelling = Not scaling



Self-supervised pre-training for image and vision

- Language: BERT (Kenton et al, 2019)

BERT's Performance - SWAG (Situations With Adversarial Generations)

System	Dev	Test
BERT$_{LARGE}$	**86.6**	**86.3**
Human (expert)	-	85.0
OpenAI GPT	-	78
ESIM+GloVe	51.9	52.7
ESIM+ELMo	59.1	59.2

✓ Better sample efficiency in downstream tasks
✓ Better multi-task performance
✓ Better transfer performance
✓ Better robustness
✓ ……

- Vision: CLIP (Radford et al, 2021)

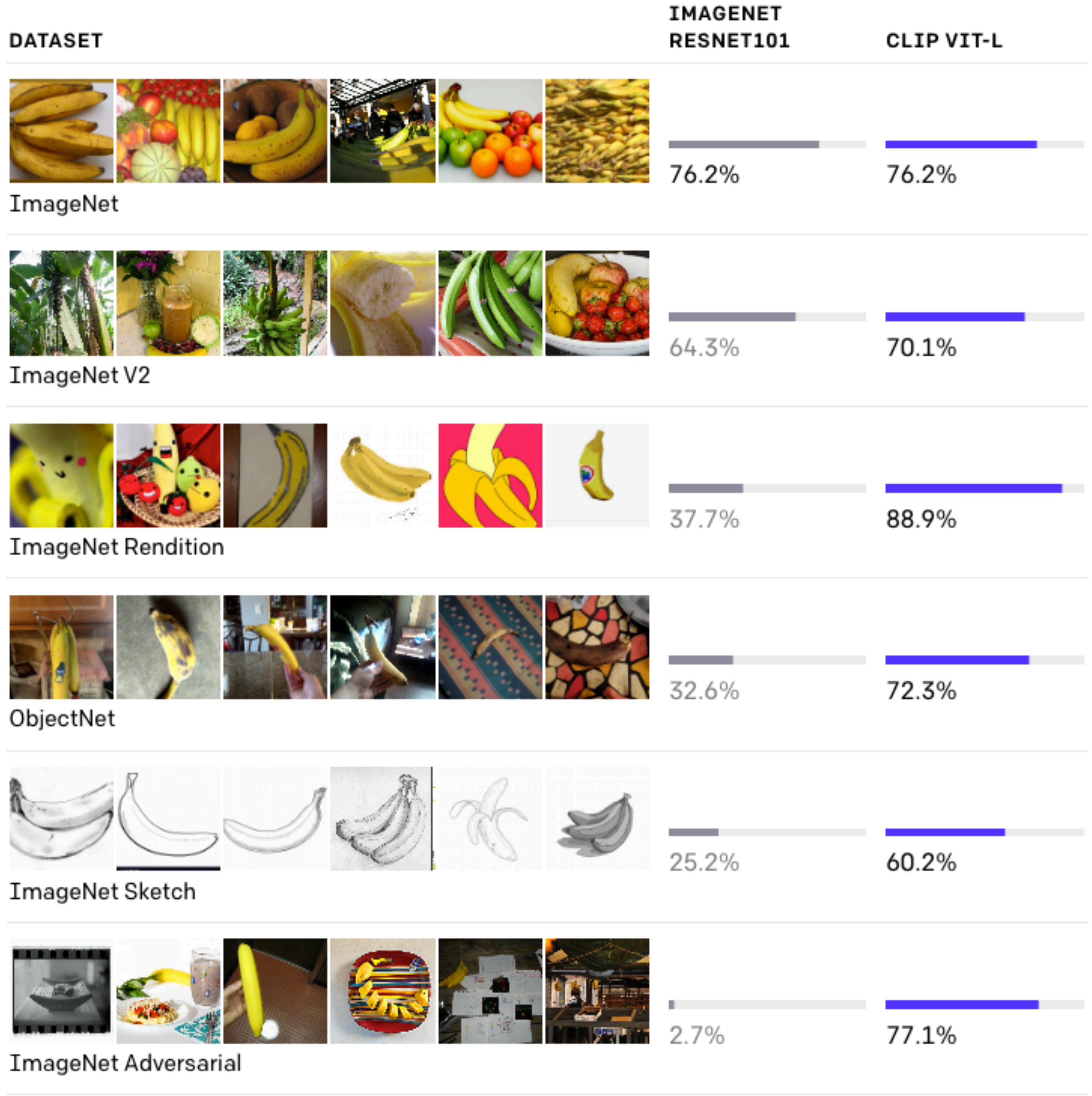
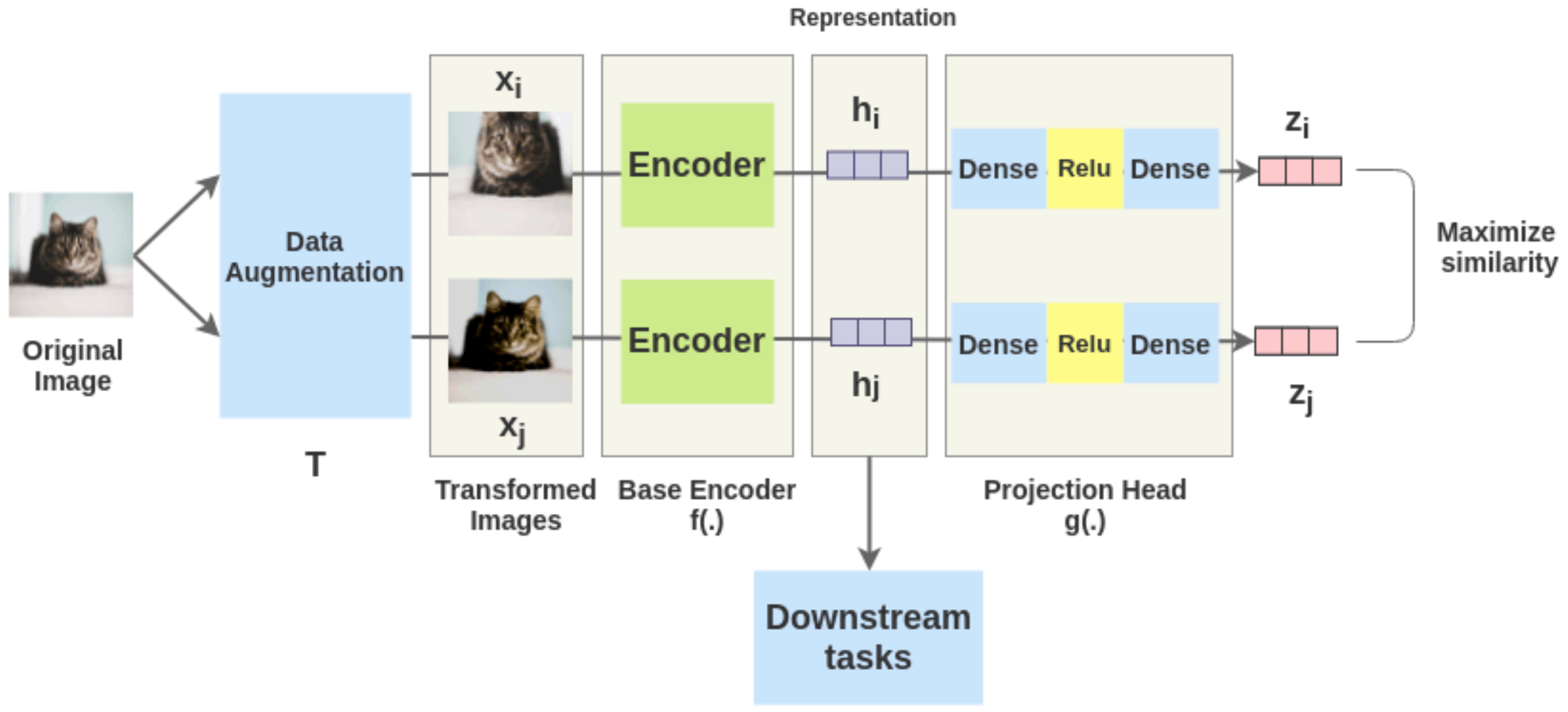


Image source: https://huggingface.co/blog/bert-101
https://openai.com/blog/clip/

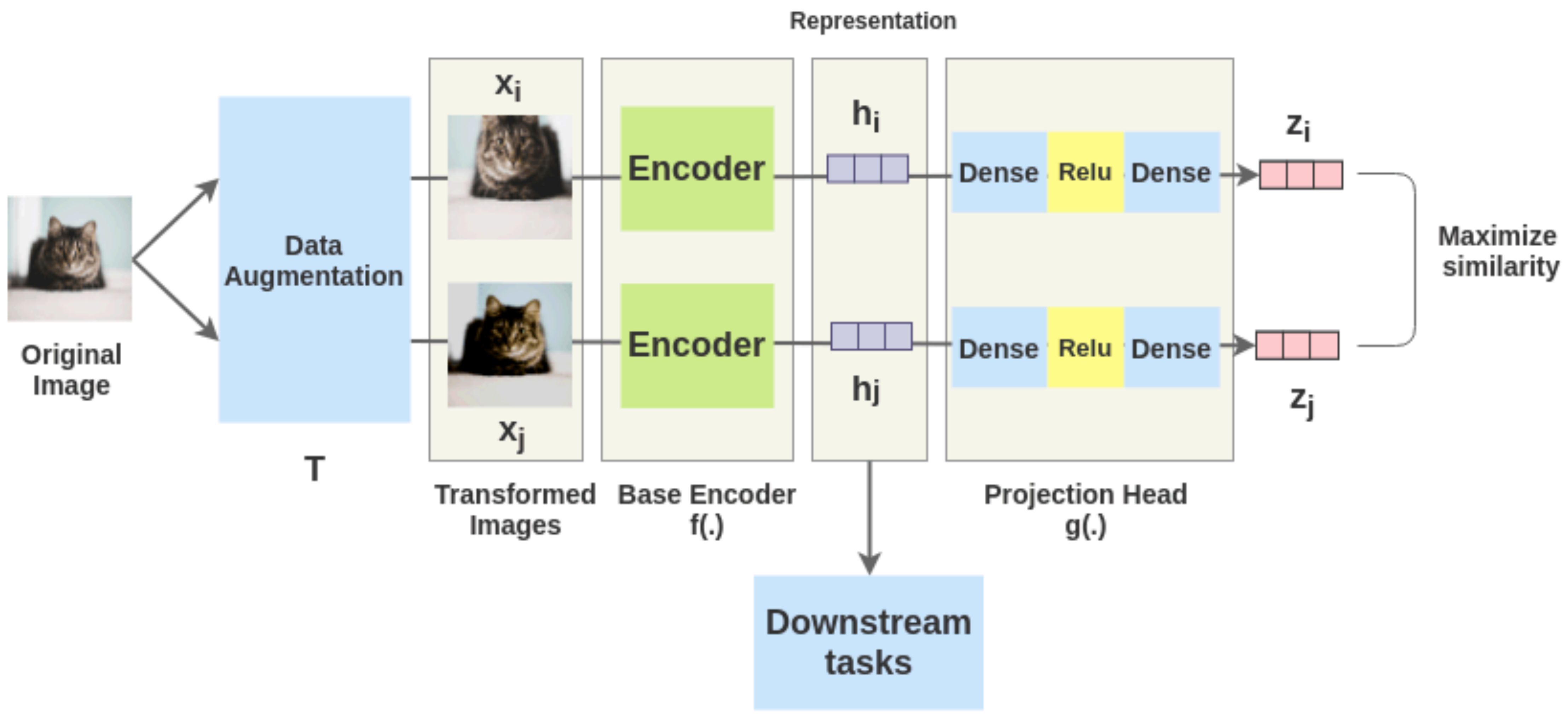
SimCLR: contrastive learning for image representations



Maximize the agreement between different augmented views of the same data

Image source: https://amitness.com/2020/03/illustrated-simclr/

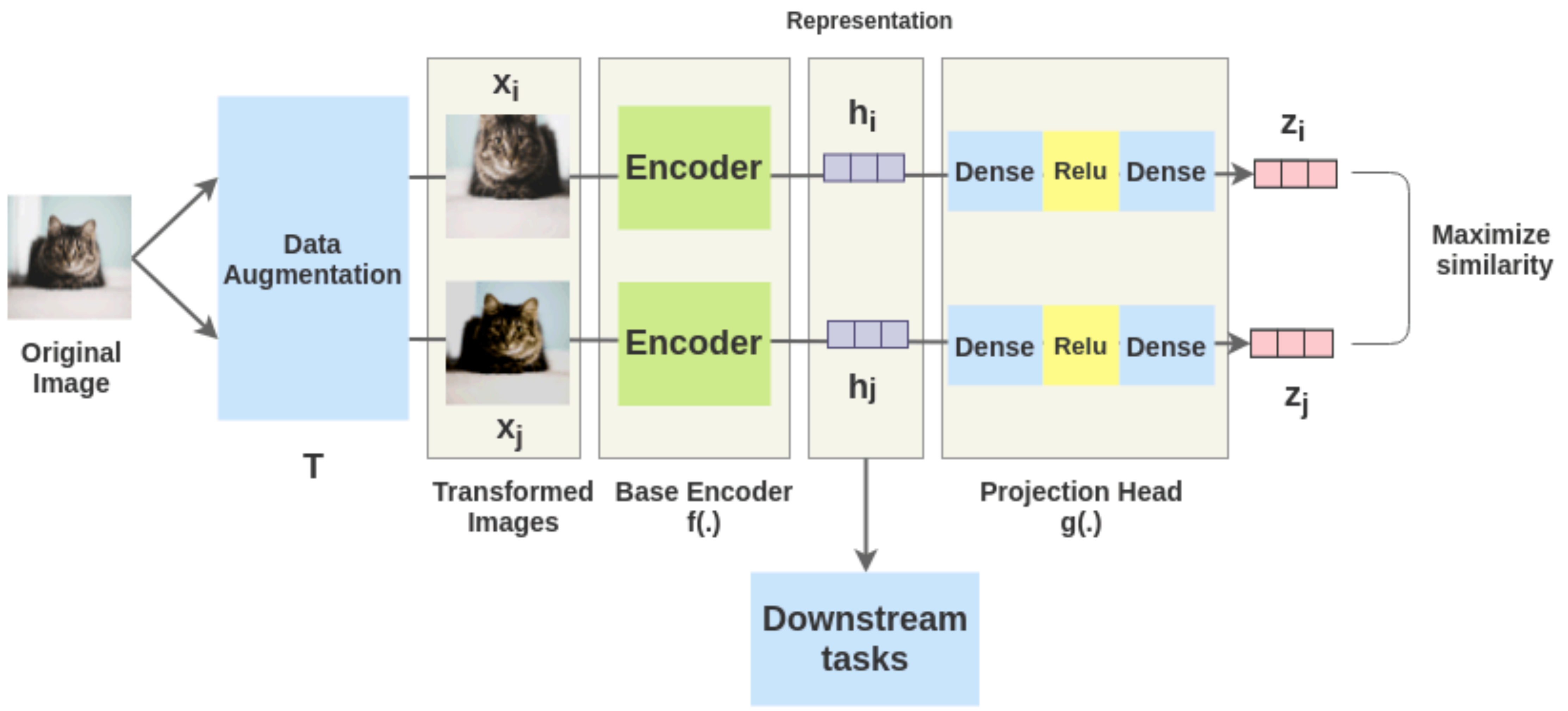
SimCLR: contrastive learning for image representations



Outperforming AlexNet on ImageNet using only 1% of the labels

Image source: https://amitness.com/2020/03/illustrated-simclr/

SimCLR: contrastive learning for image representations

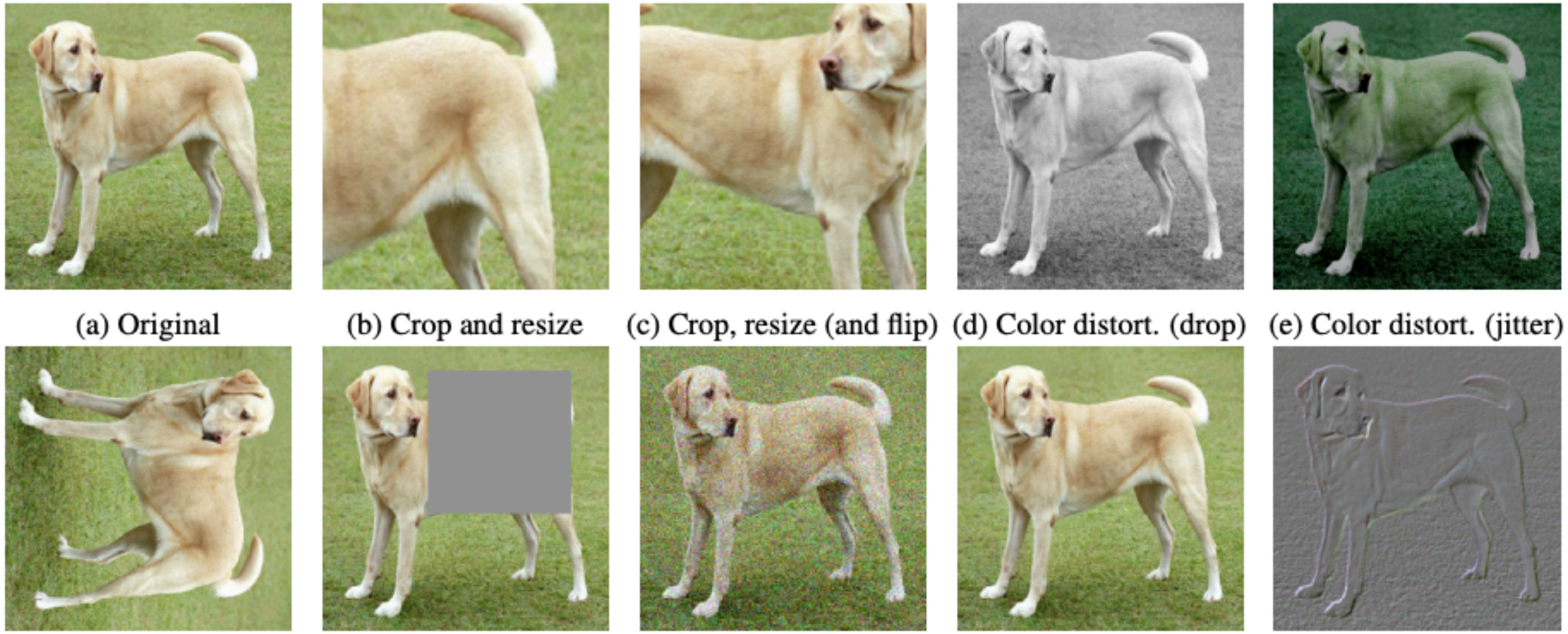


Outperforming AlexNet on ImageNet using only 1% of the labels

SimCLR for SAT representations?

Image source: https://amitness.com/2020/03/illustrated-simclr/

How to design augmentations? - Image



(a) Original (b) Crop and resize (c) Crop, resize (and flip) (d) Color distort. (drop) (e) Color distort. (jitter)

(f) Rotate {90°, 180°, 270°} (g) Cutout (h) Gaussian noise (i) Gaussian blur (j) Sobel filtering

Image source: https://arxiv.org/pdf/2002.05709.pdf

How to design augmentations? - SAT

- Requirements: the augmentations should
    - preserve labels (satisfiability): SAT => SAT, UNSAT => UNSAT
    - efficient to compute

How to design augmentations? - SAT

- Requirements: the augmentations should
  - preserve labels (satisfiability): SAT => SAT, UNSAT => UNSAT
  - efficient to compute

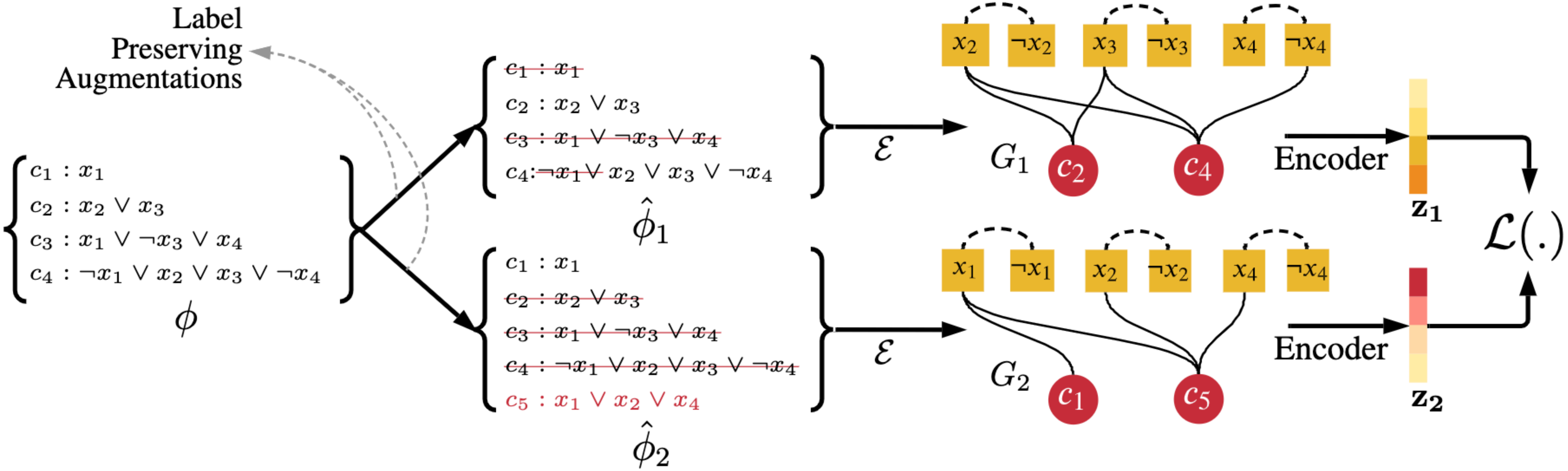
- The algorithms used in preprocessing components of SAT solvers are the perfect candidate.

Label-preserving augmentations (LPAs)

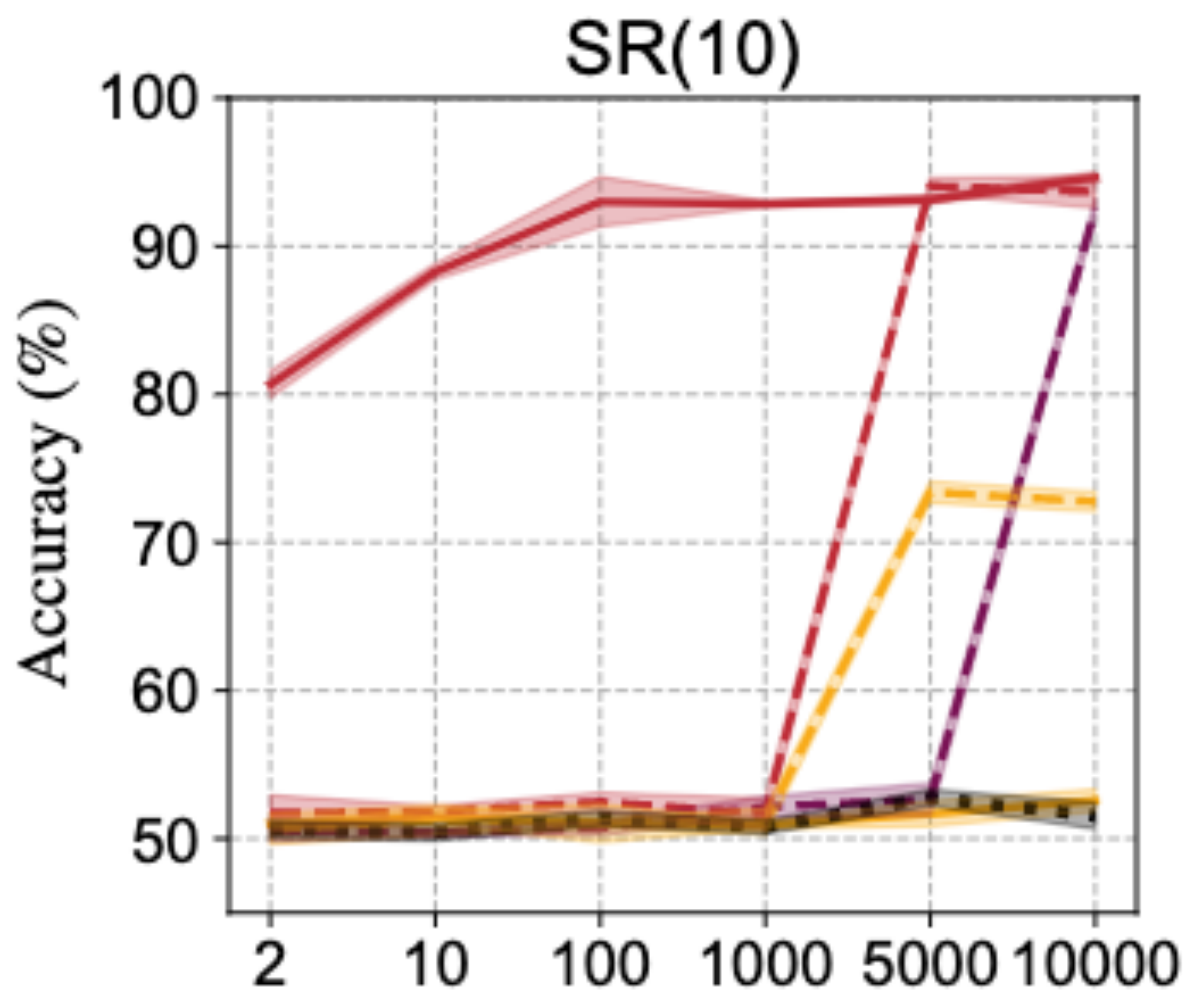
- Unit propagation (UP)
- Clause resolution (CR)
- Variable elimination (VE)
- Subsumed clause elimination (SC)
- ……

Original	UP
$c_1 : x_1$	$\cancel{x_1}$
$c_2 : x_2 \lor x_3$	$x_2 \lor x_3$
$c_3 : x_1 \lor \neg x_3 \lor x_4$	$\cancel{x_1 \lor \neg x_3 \lor x_4}$
$c_4 : \neg x_1 \lor x_2 \lor x_3 \lor \neg x_4$	$\cancel{\neg x_1 \lor} x_2 \lor x_3 \lor \neg x_4$
AU	**SC**
$\neg x_5$	
$x_5 \lor x_1$	x_1
$x_2 \lor x_3$	$x_2 \lor x_3$
$x_1 \lor \neg x_3 \lor x_4$	$x_1 \lor \neg x_3 \lor x_4$
$\neg x_1 \lor x_2 \lor x_3 \lor \neg x_4$	$\cancel{\neg x_1 \lor x_2 \lor x_3 \lor \neg x_4}$
$\neg x_5 \lor x_1 \lor \neg x_2 \lor x_3$	
CR	**VE**
x_1	x_1
$x_2 \lor x_3$	$\cancel{x_2 \lor x_3}$
$x_1 \lor \neg x_3 \lor x_4$	$\cancel{x_1 \lor \neg x_3 \lor x_4}$
$\neg x_1 \lor x_2 \lor x_3 \lor \neg x_4$	$\cancel{\neg x_1 \lor x_2 \lor x_3 \lor \neg x_4}$
$x_1 \lor x_2 \lor x_4$	$x_1 \lor x_2 \lor x_4$

Our framework



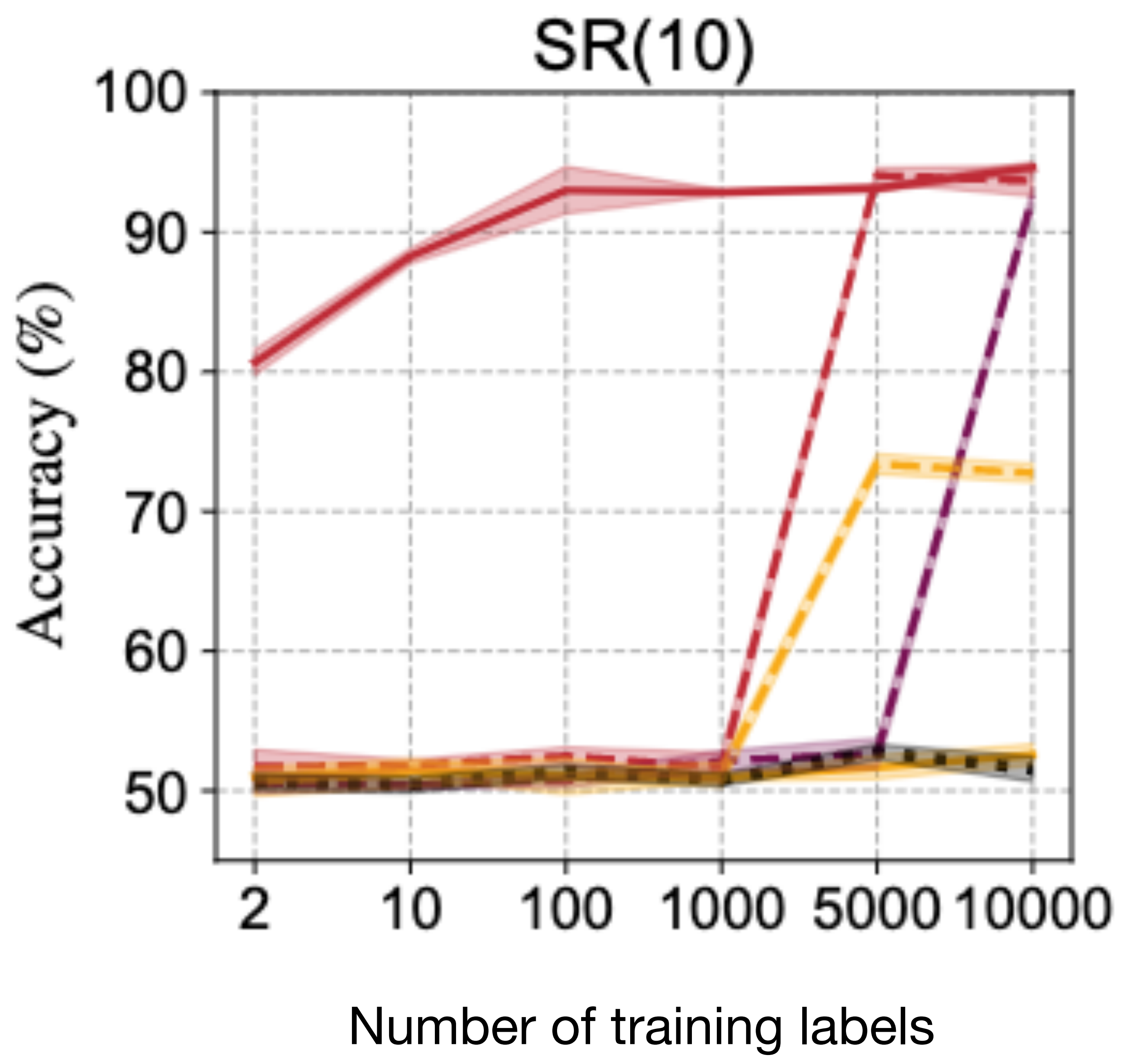
Results: pre-training improves sample efficiency by 100x



Number of training labels

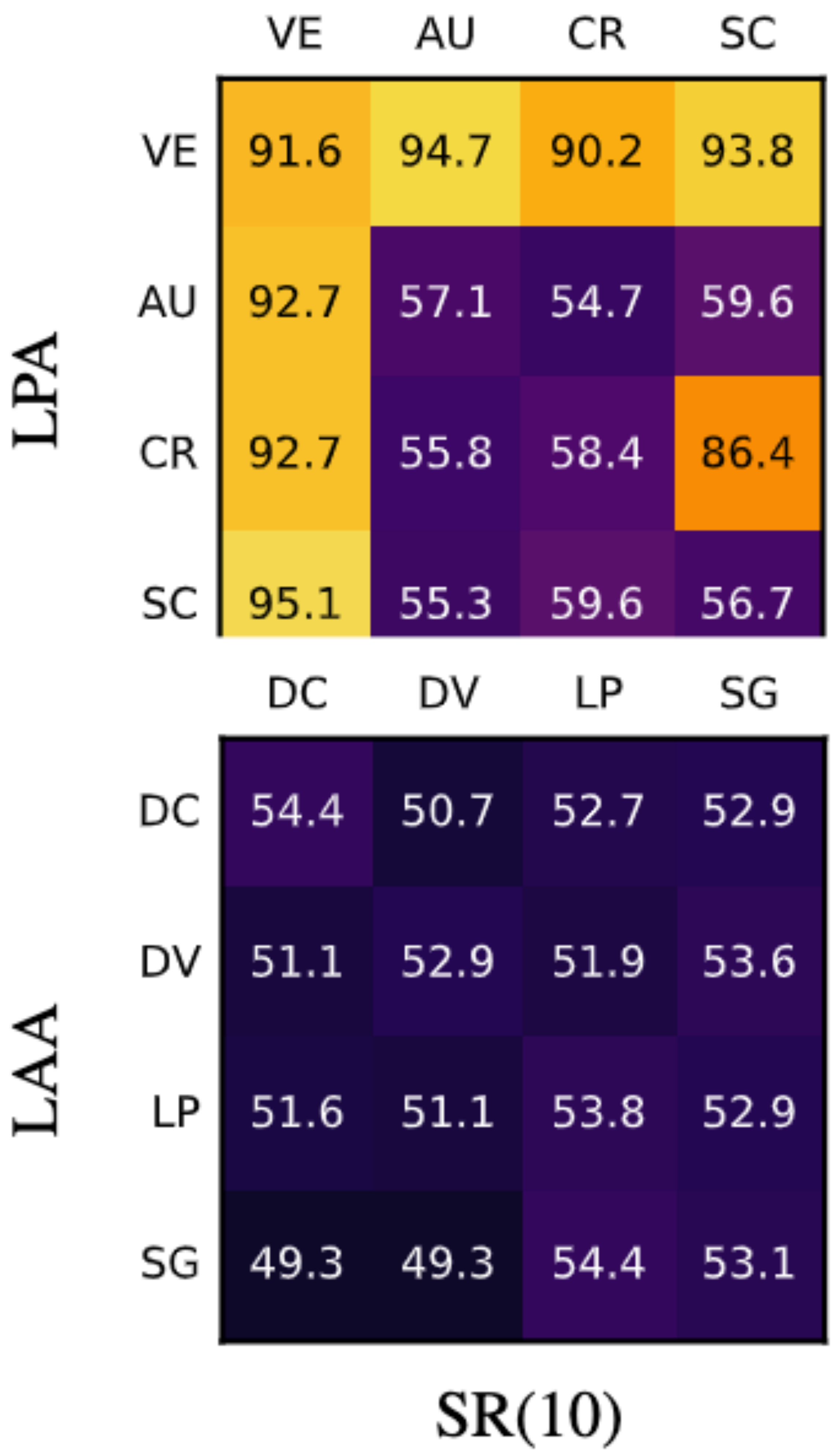
- Red line: Our pre-training + linear evaluation
- Purple line: Fully-supervised NeuroSAT

Results: pre-training improves sample efficiency by 100x



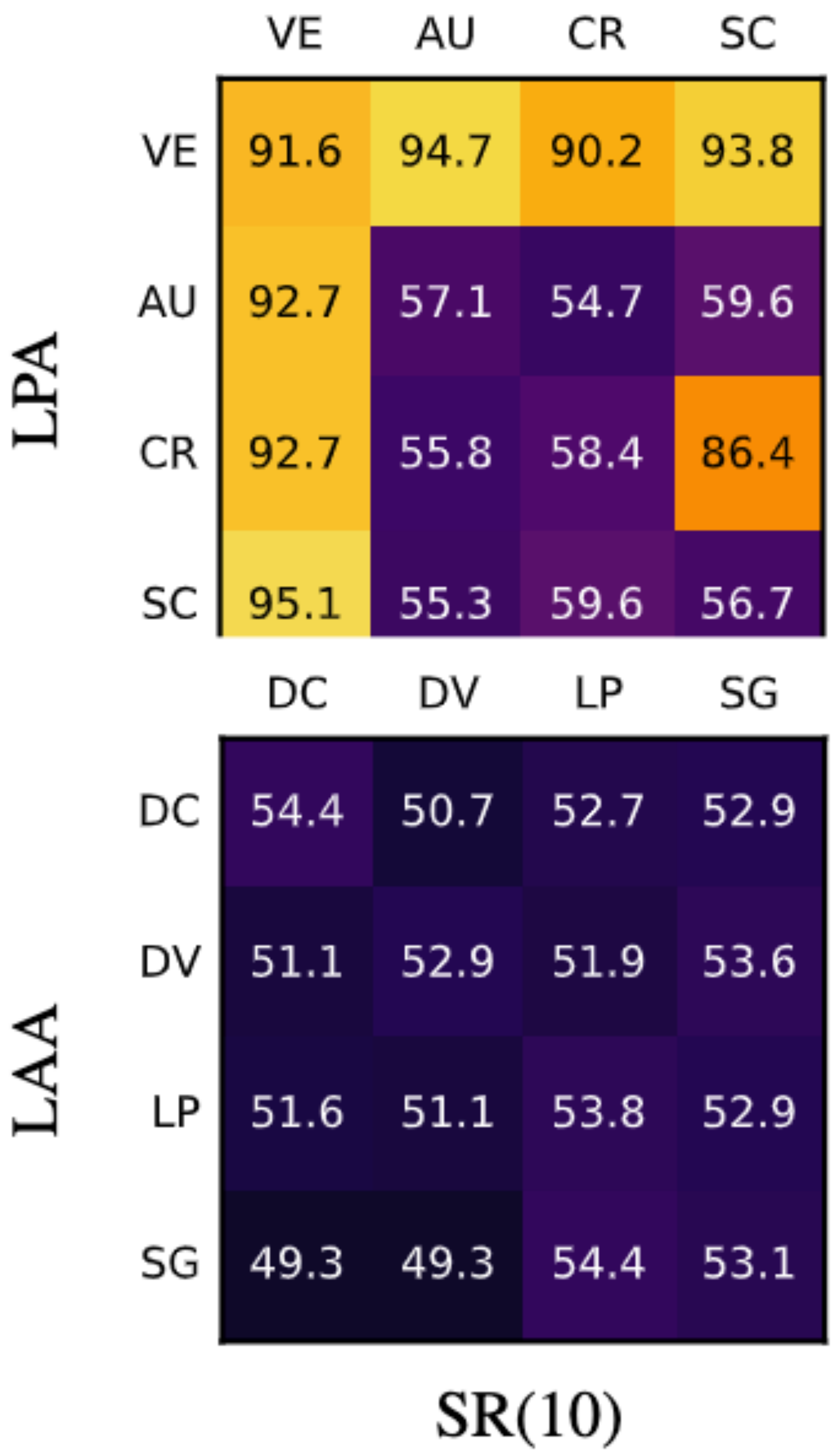
- Red line: Our pre-training + linear evaluation
- Purple line: Fully-supervised NeuroSAT
- Red with 100 labels > Purple with 10000 labels
- More datasets and settings? Check out our paper

Label-preserving augmentations are necessary



- Label-agnostic augmentations (LAAs):
  - Node dropping/adding
  - Edge perturbations
  - Subgraph

Label-preserving augmentations are necessary



- Label-agnostic augmentations (LAAs):
  - Node dropping/adding
  - Edge perturbations
  - Subgraph

- Best LPA (95.1) >> Best LAA (54.4)

Thanks!

Chat with us at our poster session: 6- 8pm, Hall E #403