



# Local Dense Logit Relations for Enhanced Knowledge Distillation

Liuchi Xu<sup>1</sup>, Kang Liu<sup>2</sup>, Jinshuai Liu<sup>1</sup>, Lu Wang<sup>1(✉)</sup>, Lisheng Xu<sup>1</sup>, Jun Cheng<sup>3(✉)</sup>

<sup>1</sup> Northeastern University, Shenyang, China

<sup>2</sup> South China Normal University, China

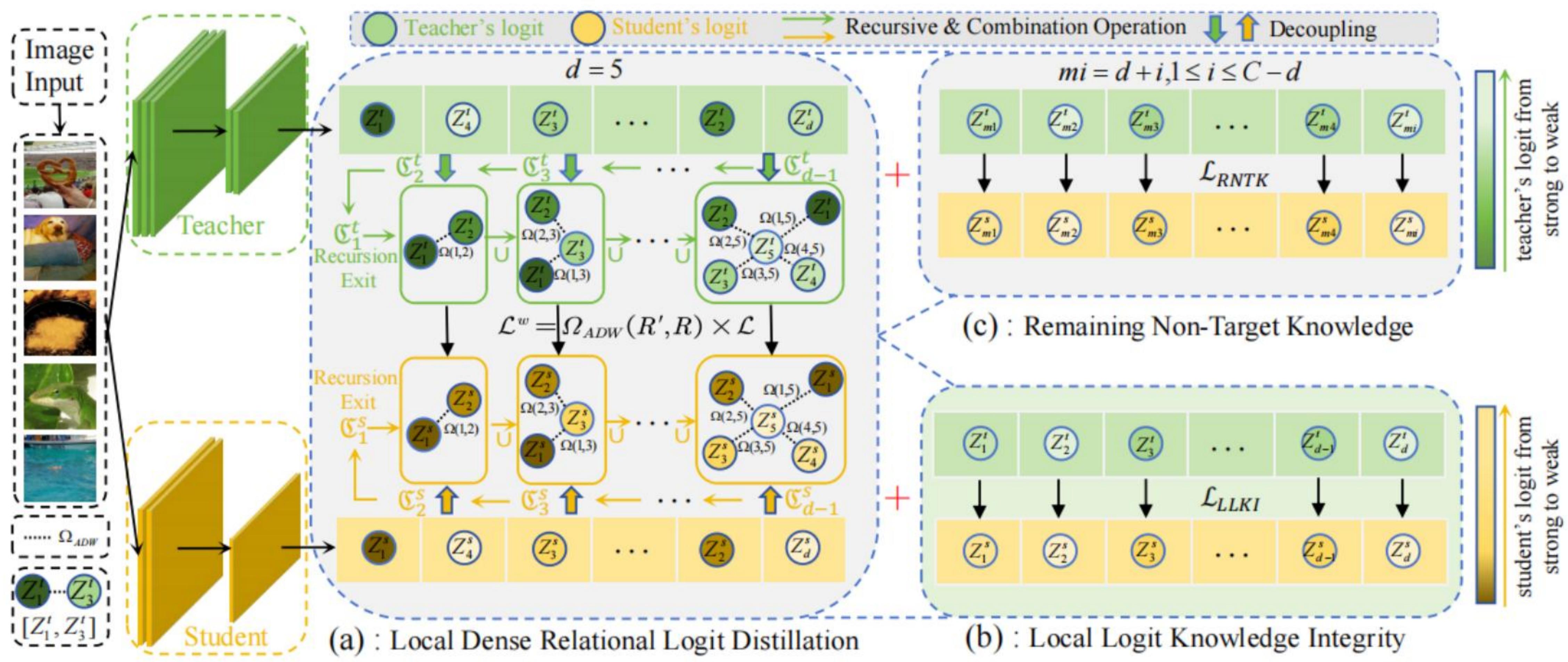
<sup>3</sup> Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China



## Contributions

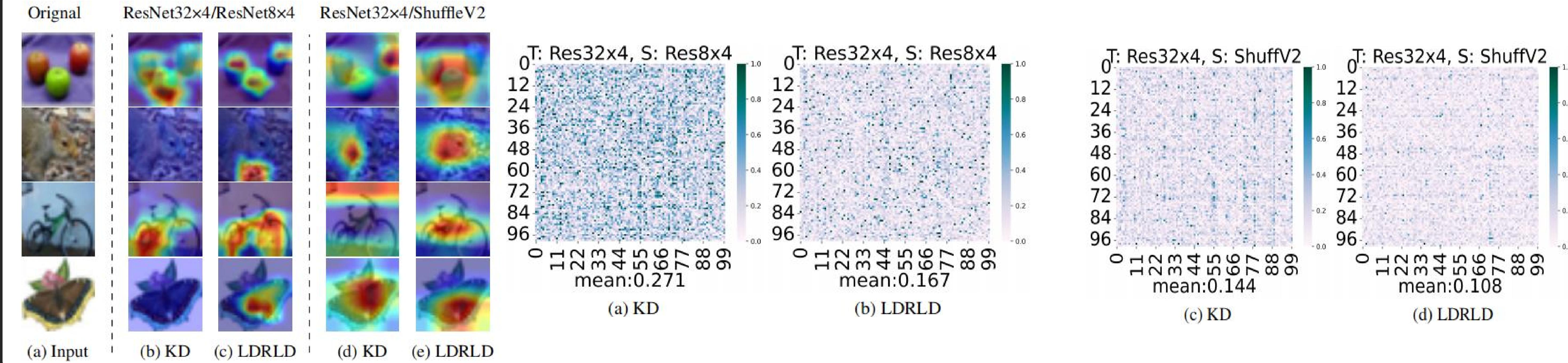
- ◆ We propose Local Dense Relational Logit Distillation (LDRRLD), a novel method that captures fine-grained logit relationships more effectively and enhances inter-class discriminability.
- ◆ We introduce the Adaptive Decay Weight (ADW) strategy, which comprises Inverse Rank Weighting (IRW) and Exponential Rank Decay (ERD). ADW can dynamically adjust the weights of ranked category pairs, thus enabling the student to more effectively optimize the classification of challenging categories.
- ◆ Extensive experimental results on diverse datasets, including CIFAR-100, ImageNet-1K, and Tiny-ImageNet, consistently show that our method outperforms existing state-of-the-art logit-based KD methods, and justify its ability to capture and transfer critical inter-class relationships.

## Proposed LDRRLD



Overview of the proposed LDRRLD framework, which includes the three key loss functions

## Visualization Experiment



## Experiment Results

ResNet34 (teacher): 73.31% Top-1, 91.42% Top-5 accuracy. ResNet18 (student): 69.75% Top-1, 89.07% Top-5 accuracy.															
Features	AT[27]	CRD[62]	ReviewKD[3]	FCFD[37]	CAT-KD[11]	Logits	KD[18]	CTKD[32]	DKD[84]	LSKD[60]	SDD[66]	WTMM[87]	TeKAP[19]	LDRRLD	Δ
Top-1	70.69	71.17	71.61	<b>72.24</b>	71.26	Top-1	70.66	71.32	71.70	71.42	71.44	<b>72.19</b>	71.35*	71.88	<b>+1.22</b>
Top-5	90.01	90.13	90.51	<b>90.74</b>	90.45	Top-5	89.88	90.27	90.31	90.29	90.05	-	90.54*	<b>90.58</b>	<b>+0.70</b>
ResNet50 (teacher): 76.16% Top-1, 92.87% Top-5 accuracy. MobileNetV1 (student): 68.87% Top-1, 88.76% Top-5 accuracy.															
Features	AT[27]	CRD[62]	ReviewKD[3]	FCFD[37]	CAT-KD[11]	Logits	KD[18]	IPWD[43]	DKD[84]	LSKD[60]	SDD[66]	WTMM[87]	TeKAP[19]	LDRRLD	Δ
Top-1	70.18	71.32	72.56	<b>73.37</b>	72.24	Top-1	70.49	72.65	72.05	72.18	72.24	73.09	72.87*	<b>73.12</b>	<b>+2.63</b>
Top-5	89.68	90.41	91.00	<b>91.35</b>	91.13	Top-5	89.92	91.08	91.05	90.80	90.71	-	91.05*	<b>91.43</b>	<b>+1.51</b>

Table 3. Evaluate the top-1 and top-5 accuracy (%) of student using same and different architectures on the ImageNet-1K validation set.

Teacher	Student	From Scratch		Feature-based						Logit-based					
		T:Accuracy	S:Accuracy	FitNet[51]	RKD[44]	CRD[62]	FOFA[34]	KD[18]	DKD[84]	DIST[21]	OFA[14]	TeKAP[19]	LDRRLD	Δ	
Swin-T	ResNet18	89.26	74.01	78.87	74.11	77.63	81.22	78.74	80.26	77.75	80.54	81.38*	<b>82.17</b>	<b>+3.43</b>	
ViT-S	ResNet18	92.04	74.01	77.71	73.72	76.60	80.11	77.26	78.10	76.49	80.15	79.06*	<b>80.36</b>	<b>+3.10</b>	
Mixer-B/16	ResNet18	87.29	74.01	77.15	73.75	76.42	80.07	77.79	78.67	76.36	79.39	80.05*	<b>80.69</b>	<b>+2.90</b>	
Swin-T	MobileNetV2	89.26	73.68	74.28	69.00	79.80	78.78	74.68	71.07	72.89	80.98	80.23*	<b>81.64</b>	<b>+4.96</b>	
ViT-S	MobileNetV2	92.04	73.68	73.54	68.46	78.14	78.87	72.77	69.80	72.54	78.45	78.41*	<b>79.21</b>	<b>+6.44</b>	
Mixer-B/16	MobileNetV2	87.29	73.68	73.78	68.95	78.15	78.62	73.33	70.20	73.26	78.78	79.89*	<b>80.64</b>	<b>+7.31</b>	
ConvNeX-T	DeiT-T	88.41	68.00	60.78	69.79	65.94	79.59	72.99	74.60	73.55	75.76	76.32*	<b>77.46</b>	<b>+4.47</b>	
Mixer-B/16	DeiT-T	87.29	68.00	71.05	69.89	65.35	74.66	71.36	73.44	71.67	73.90	74.83*	<b>75.31</b>	<b>+3.95</b>	
ConvNeX-T	Swin-P	88.41	72.63	24.06	71.73	67.09	80.74	76.44	76.80	76.41	78.32	79.18*	<b>80.71</b>	<b>+4.27</b>	
Mixer-B/16	Swin-P	87.29	72.63	75.20	70.82	67.03	78.44	75.93	76.39	75.85	78.93	78.97*	<b>80.52</b>	<b>+4.59</b>	
ConvNeX-T	ResMLP-S12	88.41	66.56	45.47	65.82	63.35	83.50	72.25	73.22	71.93	75.21	81.14*	79.28	<b>+7.03</b>	
Swin-T	ResMLP-S12	89.26	66.56	63.12	64.66	61.72	80.94	71.89	72.82	11.05	73.58	80.22*	<b>80.54</b>	<b>+8.65</b>	

Table 4. Evaluation of the top-1 accuracy (%) of student using ViT-based heterogeneous architectures on the CIFAR100 dataset.

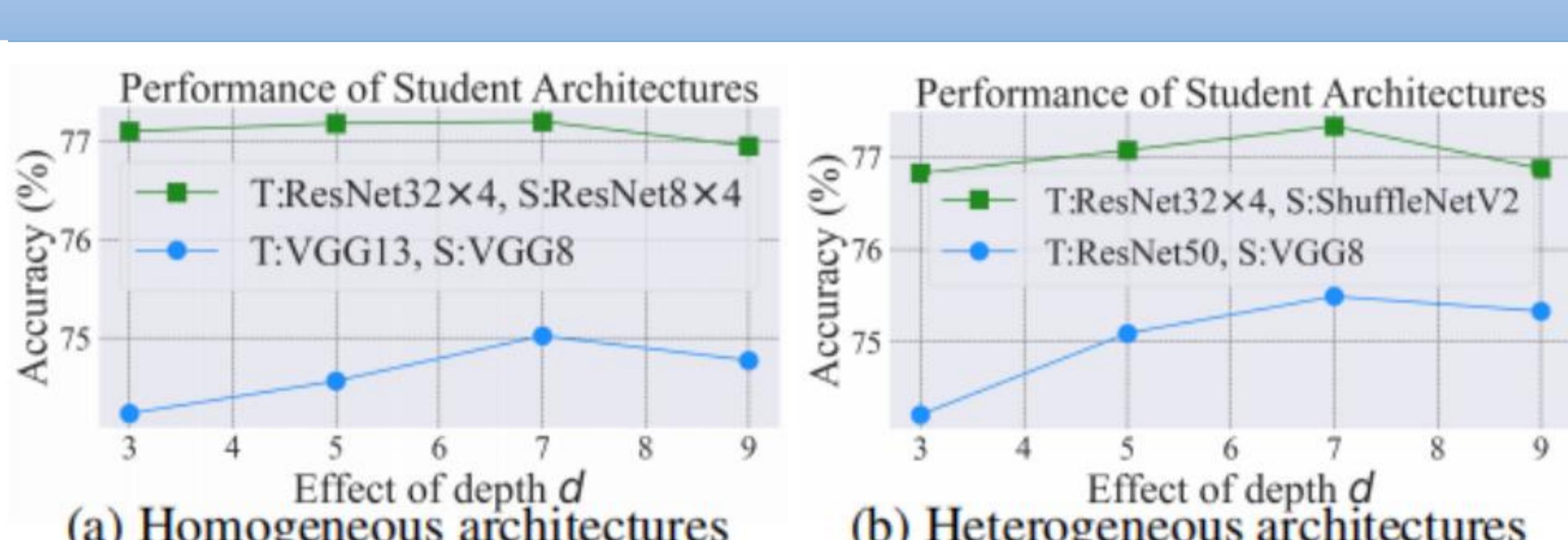


Figure 3. Impact of the depth  $d$  on the performance of the student on CIFAR-100.

$\mathcal{L}_{Local}$	VGG13		ResNet50		ResNet32x4	
$\mathcal{L}$	$\Omega_{ADW}$	$\mathcal{L}_{LLKI}$	VGG8	MobileNetV2	MobileNetV2	ShuffleNetV2
-	-	-	70.50	64.60	64.60	71.82
✓		✓	74.25 (+3.75)	68.95 (+4.35)	69.23 (+4.63)	76.92 (+5.10)
✓	✓	✓	<b>74.46 (+3.96)</b>	<b>69.27 (+4.67)</b>	<b>69.52 (+4.92)</b>	<b>77.07 (+5.25)</b>