Supplementary: Jumpstarting Surgical Computer Vision

Implementation Details

Parameter	Value	Parameter	Value
Method	TeCNO [3]	Class Weighting	Median Frequency Balancing
Stage 1 Learning Rate	3e-3	Stage 1 Optimizer	SGD with LARC [6]
Stage 2 Learning Rate	3e-3	Stage 2 Optimizer	Adam [4]
GPUs	4 NVidia V100s		

Table 1. Various design choices and implementation hardware pertaining to all downstream runs for all phase recognition experiments (Cholec80, AutoLaparo).

Parameter	Value	Parameter	Value
Method	Linear Finetuning	Class Weighting	Inverse Frequency Balancing
Learning Rate	1e-5	Optimizer	AdamW [5]
GPUs	1 NVidia V100s		

Table 2. Various design choices and implementation hardware pertaining to all downstream runs for all CVS experiments (EndoScapes-201CVS).

Parameter	Value	Parameter	Value
Architecture	ResNet-50	Queue size	65536
Method	MoCo v2 [1]	Decay parameter (λ)	0.999
Projection Head	3 layer Multi-Layer Perceptron	Output Dimension Size	4096
GPU	4 NVidia V100s	Optimizer	LARC
Multi-Crop size	224x224	Number of crops	2
Batch Size	256	Epochs	300
Sampling Rate	5	Base initialization	Imagenet Sup.

Table 3. SSL hyperparameters specific to the SSL pre-training implementation of the MoCo v2 pre-training runs. Compute of $\sim$ 42 GPU hours was required per pre-training run.

	Name	Source	Parameters
Color	Sharpness	RA	$prob = 0.5, M = 8.0, M_\sigma = 0.5$
	Brightness	RA	$prob = 0.5, M = 8.0, M_\sigma = 0.5$
	Contrast	RA	$prob = 0.5, M = 8.0, M_\sigma = 0.5$
	Color	RA	$prob = 0.5, M = 8.0, M_\sigma = 0.5$
	Auto-contrast	RA	$prob = 0.5, M = 8.0, M_\sigma = 0.5$
	Equalize	RA	$prob = 0.5, M = 8.0, M_\sigma = 0.5$
	Random-erasing	torchvision	$prob = 0.8, scale = [0.02, 0.1]$
Strong Color	Posterize	RA	$prob = 0.5, M = 8.0, M_\sigma = 0.5$
	Solarize	RA	$prob = 0.5, M = 8.0, M_\sigma = 0.5$
	invertion	RA	$prob = 0.5, M = 8.0, M_\sigma = 0.5$
Geometric	Rotate	RA	$prob = 0.5, M = 8.0, M_\sigma = 0.5$
	Translate-x	RA	$prob = 0.5, M = 8.0, M_\sigma = 0.5$
	Translate-y	RA	$prob = 0.5, M = 8.0, M_\sigma = 0.5$
	Shear-x	RA	$prob = 0.5, M = 8.0, M_\sigma = 0.5$
	Shear-y	RA	$prob = 0.5, M = 8.0, M_\sigma = 0.5$
	Horizontal-flip	torchvision	$prob = 0.5$
Multi Crop	MC2: multi-crop 2	torchvision	$224x224 = 2$, $scale = [0.5, 1]$

Table 4. Different augmentations used in the SSL pre-training experiments. The Source "RA" uses the RandAugment [2] implementations that randomly selects two augmentations from the list, applying the augmentation with a given magnitude (M), probability ($prob$), and standard deviation on the magnitude (M_σ). The source "torchvision" uses the torchvision implementation for Horizontal-flip, Random-erasing, and Multi-crop augmentations.

References

1. Chen, X., Fan, H., Girshick, R., He, K.: Improved baselines with momentum contrastive learning. arXiv preprint arXiv:2003.04297 (2020)
2. Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.: Randaugment: Practical automated data augmentation with a reduced search space. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual (2020)
3. Czempiel, T., Paschali, M., Keicher, M., Simson, W., Feussner, H., Kim, S.T., Navab, N.: Tecno: Surgical phase recognition with multi-stage temporal convolutional networks. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part III 23. pp. 343–352. Springer (2020)
4. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
5. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
6. You, Y., Gitman, I., Ginsburg, B.: Large batch training of convolutional networks. arXiv preprint arXiv:1708.03888 (2017)