Boosting Self-supervised Video-based Human Action Recognition Through Knowledge Distillation

Fernando Camarena (*fernando@camarenat.com*), Miguel Gonzalez-Mendoza, Leonardo Chang, and Neil Hernandez-Gress| Tecnologico de Monterrey, School of Engineering and Science. Mexico

Introduction

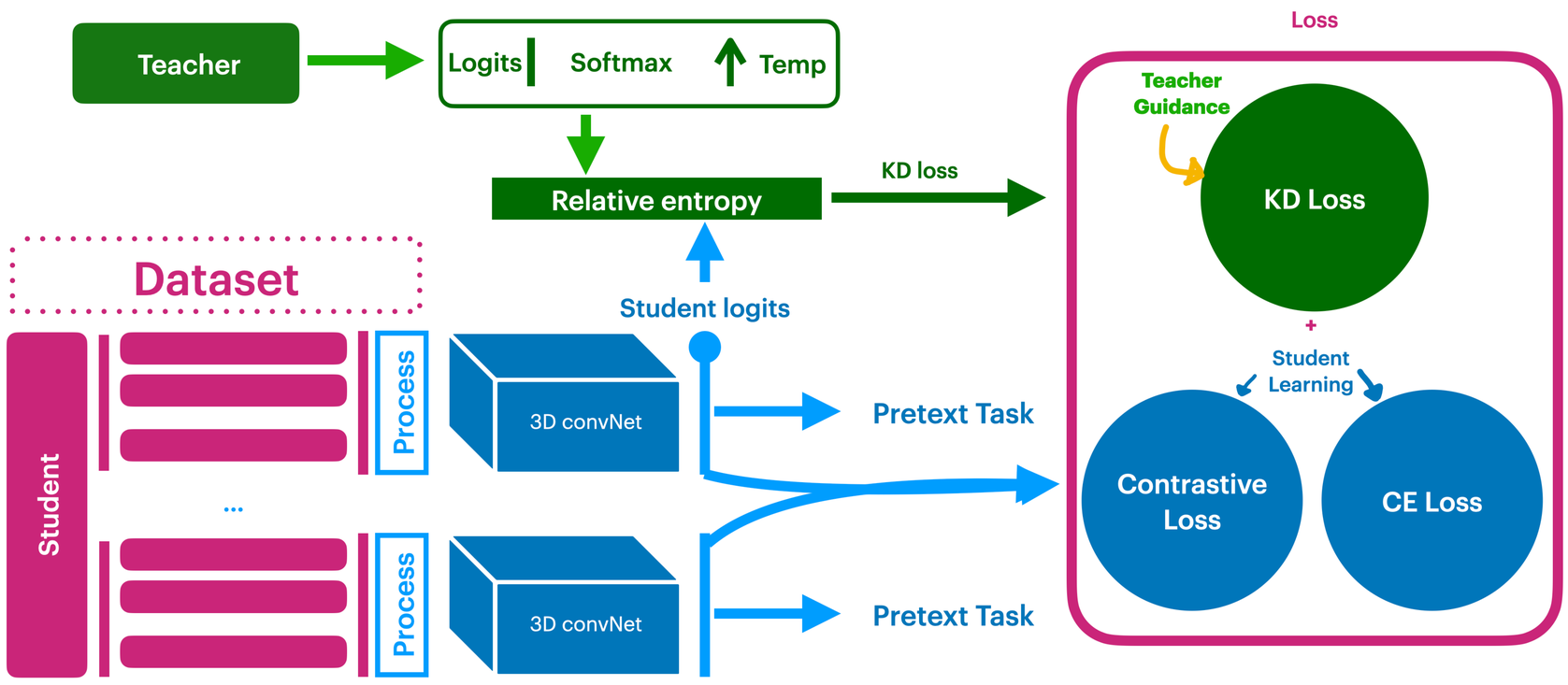
- Deep learning leads the state-of-the-art.
- Nevertheless, current methods work under a supervised methodology, requiring high-quality labels.
- Current methods like self-supervised learning use unlabeled data, but they are computationally expensive, and knowledge transfer is usually by fine-tuning.
- Fine-tuning does not enable the transfer between architecture settings.

Related Work

- **Action recognition**: understand the encoded message in a sequence of gestures.
- **Self-supervised learning** (SSL): a training method that uses a natural supervision from unlabeled data.
- **PCL**: A SSL method that combines a pretext task with contrastive learning.
- **Knowledge distillation** (KD): a novel technique for transferring knowledge that uses the pretrained model as guidance in the training algorithm.

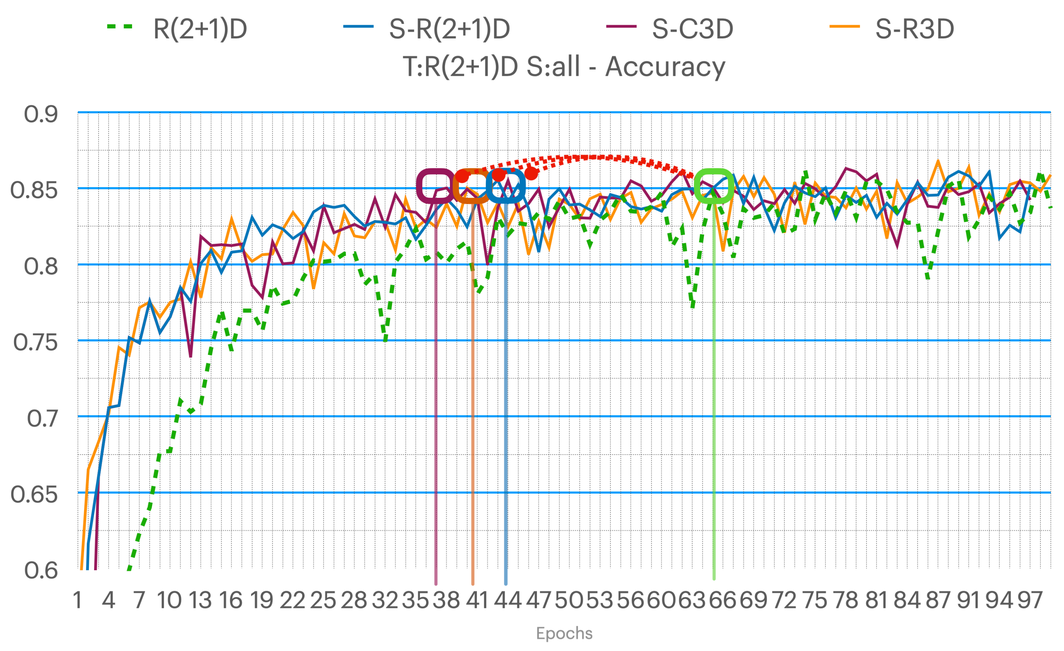
Proposed Work

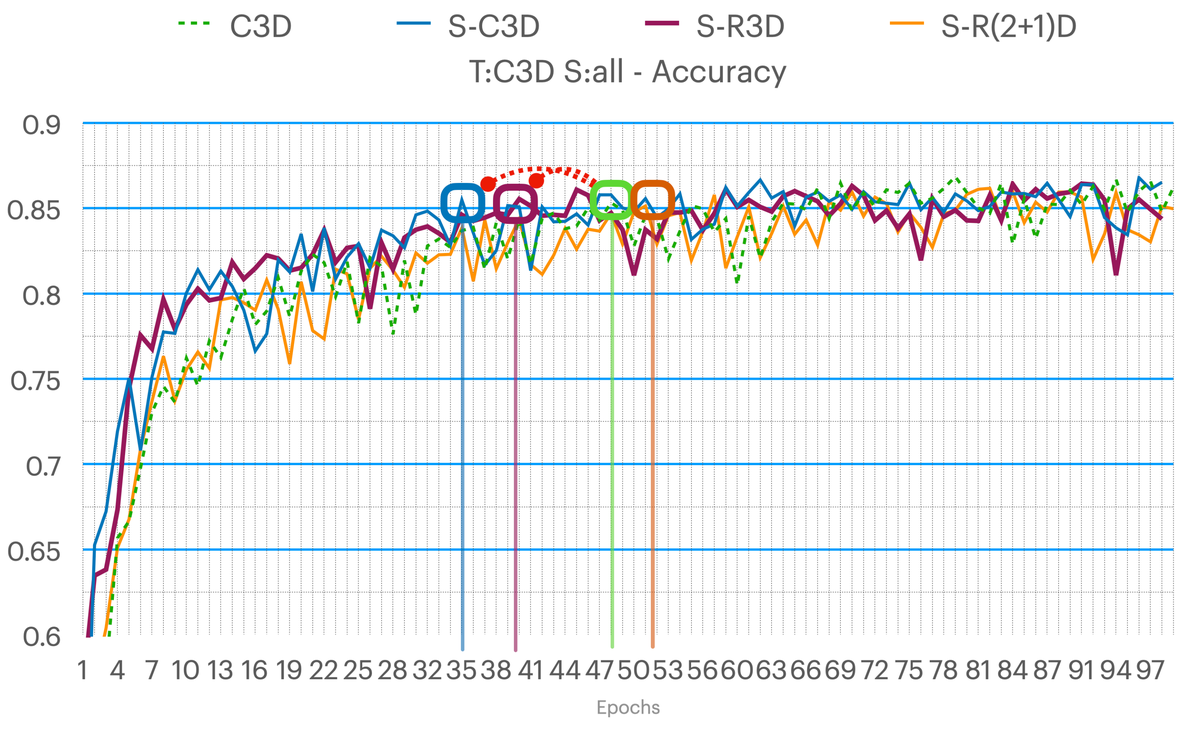
- **Intuition**: Teachers filter the knowledge and help its assimilation. KD helps to compare the probability distribution of the student and teacher networks.
- **Proposed framework**: the learning algorithm combines the relative entropy between teacher and student, the contrastive learning loss, and the pretext task loss as the loss function.
- **Relative entropy** is computed by the kulllback-leiber function between the network outputs (using a temperature to scale low probabilities values).

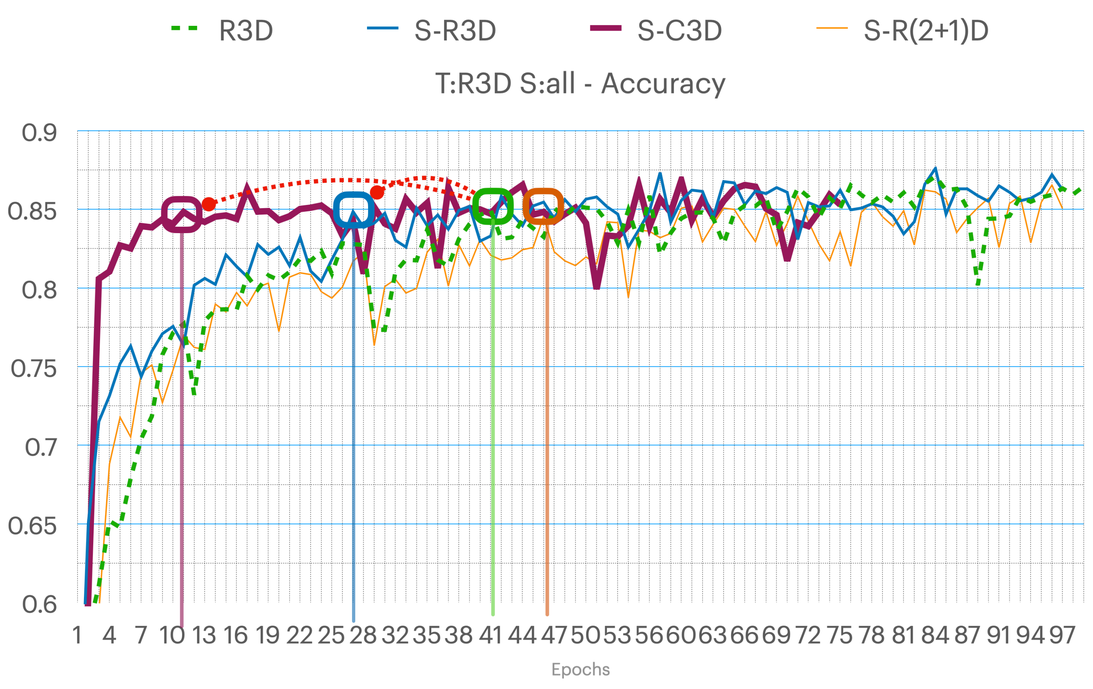


Results

- **Dataset**: UCF101 - more than 10 thousand videos.
- **Architectures** used: R3D, C3D, and R(2+1)D.
- Students **outperform** the teacher models using the same and different architectural designs.
- Students **converge faster** than the teacher models using the same and different architectures.
- Using different architectural designs boost the model's performance.







Conclusions

KD works in the video action recognition domain

KD boots SSL convergence in video action recognition

KD enables the transfer between different configuration settings

KD improves the performance of SSL algorithm