

Supplementary Material: High-Level Features for Movie Style Understanding

Robin Courant¹

Christophe Lino¹

Marc Christie²

Vicky Kalogeiton¹

¹LIX, Ecole Polytechnique, CNRS, IP Paris

²Univ Rennes, CNRS, IRISA, INRIA

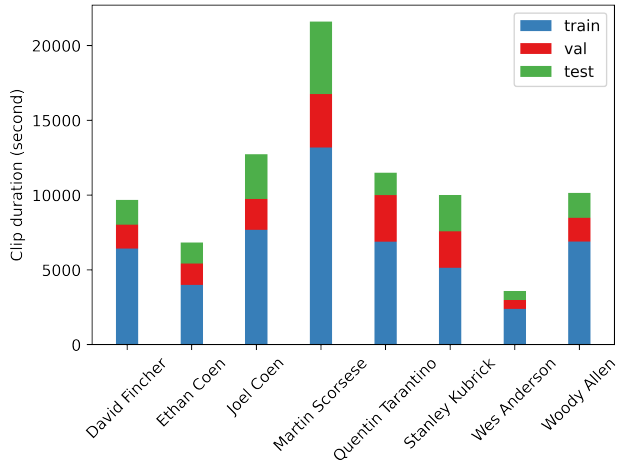


Figure 1. Distribution of the clip duration per director in CMD8.

1. Director Classification

In this section, we give more details about the dataset and the implementation that we use to tackle the director classification task.

CMD8. We choose the directors composing CMD8 on the basis of their fame. As their similarities and differences are well-known, it makes the analysis of results easier and straightforward. Figure 1 shows the clip duration distribution across directors for the train, val and test sets. Note, CMD8 is imbalanced, with an over-representation of Martin Scorsese’s movies (17 movies) and an under-representation of Wes Anderson (4 movies). However, the mean clip duration, 133.17 seconds, is approximately homogeneous across categories. The shortest and longest clips last 27 and 274 seconds, respectively.

Implementation details. Following [1], we train our model with SGD with momentum of 0.9, weight decay of 10^{-3} and a learning rate of 0.1. For data augmentation, we only apply horizontal random flipping to the training samples. Adding other types of augmentation, such as rescaling or cropping is not suitable for our task, as they deform the frames too much and remove the style information.

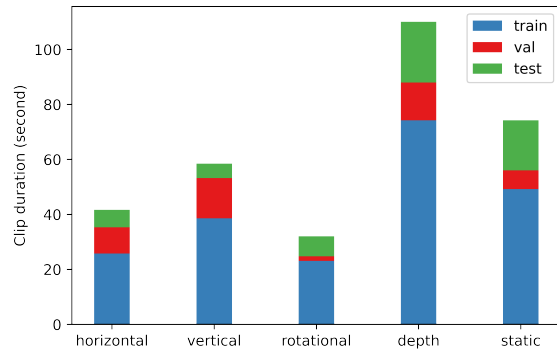


Figure 2. Distribution of the clip duration per camera motion type in MotionSet.

2. Camera Motion Detection

In this section, we give more details about the dataset and the implementation that we use to tackle the camera motion detection task.

MotionSet. Figure 2 shows the clip duration distribution across directors for the train, val and test sets. Note, MotionSet is also imbalanced, with an over-representation of depth and an under-representation of rotational shots. Moreover, the mean clip duration is not homogeneous across classes: from 2.78 sec for horizontal to 7.42 sec for static shots.

Implementation details. We use Adam with a learning rate of 10^{-4} . For data augmentation, we use random horizontal flipping, random downsampling and colour jittering.

References

- [1] Kensho Hara, Hirokatsu Kataoka, and Yutaka Satoh. Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? In *CVPR*, 2018.