



# SHOT SCALE ANALYSIS IN MOVIES BY CONVOLUTIONAL NEURAL NETWORKS

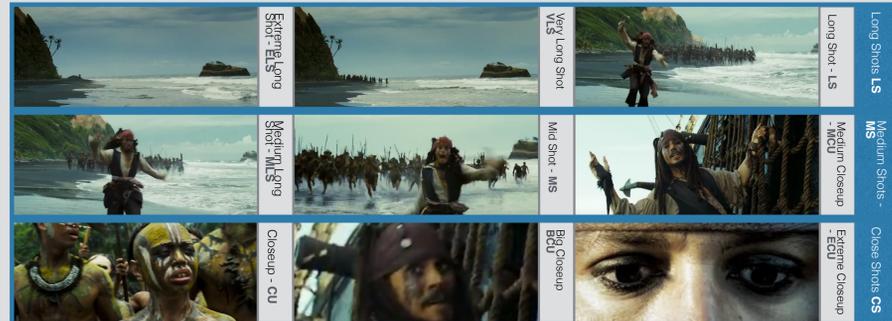


Mattia Savardi, Alberto Signoroni, Pierangelo Migliorati, Sergio Benini

Department of Information Engineering at University of Brescia (Italy)

## Introduction

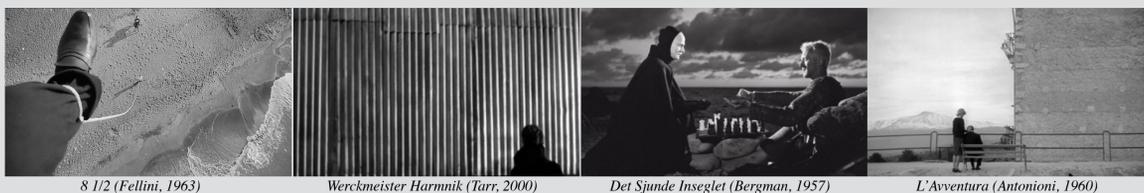
- The apparent distance of the camera from the subject of a filmed scene is called *shot scale*
- In this work we propose to use CNNs for the automatic classification of shot scale into Close-, Medium-, or Long-shots
- This allows for investigating the relationship between shot scale and the viewers' emotional involvement, for purposes such as movie recommendation, stylistic analysis, film therapy, etc.
- Shot scale induce psychological impacts on the viewers [1]. Its specific usage increases arousal, empathic care, relates memory, intensifies character liking/disliking, acting on the narrative engagement of the viewer and his/her ability of attributing mental states to movie characters
- Shot scale usage may be used as an authorial fingerprint [2] for authorship attribution
- Training and testing are performed on the filmographies by 6 authors (120 movies)
- Classification results are superior to state-of-the-art (accuracy ~94%)



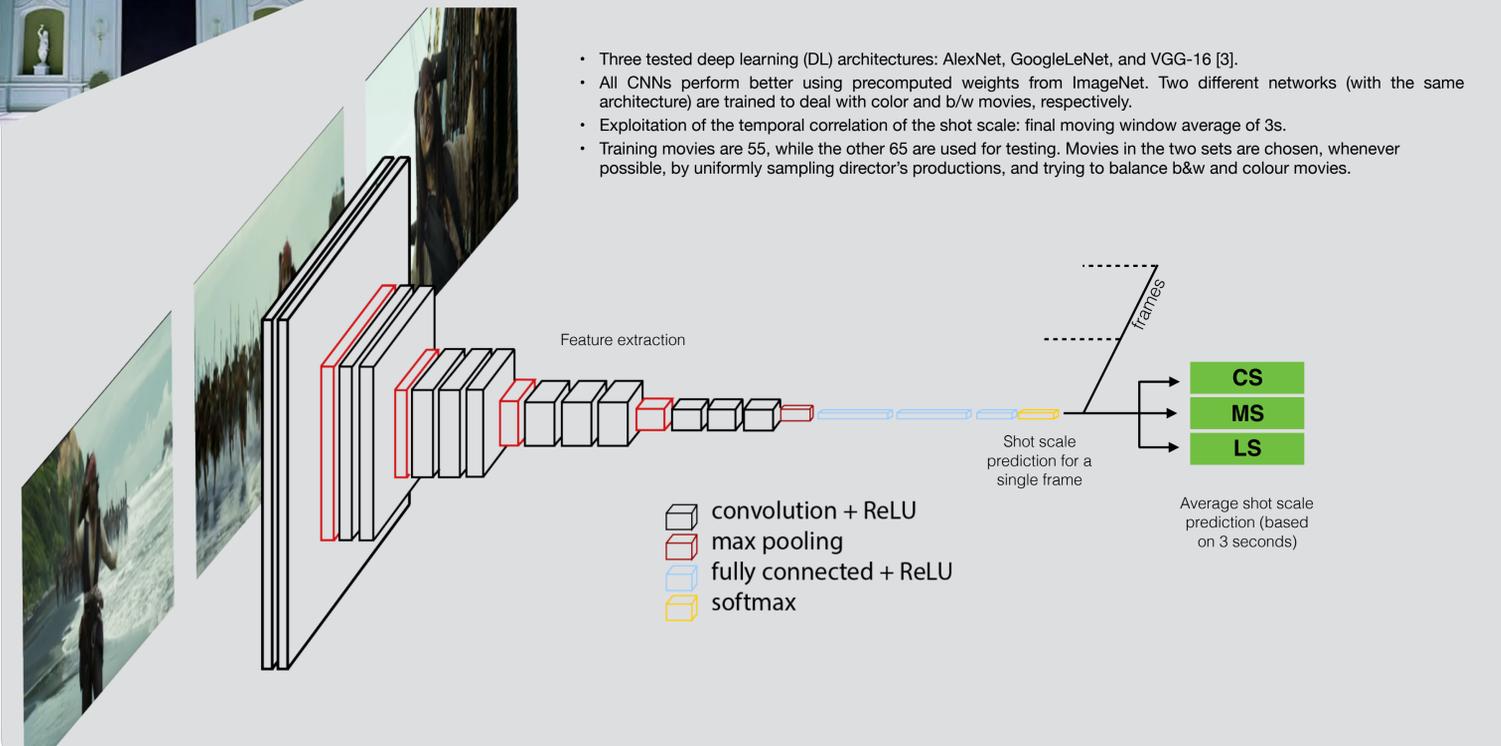
## Movie dataset

- The movie dataset includes the almost complete filmographies by six directors whose styles are consensually considered highly unique in film historiography of author cinema: Michelangelo Antonioni, Ingmar Bergman, Federico Fellini, Jean-Luc Godard, Martin Scorsese, and Bela Tarr.
- A total of 120 movies analyzed on a second base.

	CS	MS	LS
Training	Colour 57,996	25,794	9,563
	b&w 30,925	14,391	4,614
	Total 88,921	30,408	14,177
Test	Colour 64,053	22,506	11,118
	b&w 37,189	26,336	5,946
	Total 191,212	48,842	17,064
Total	280,133	79,250	31,241

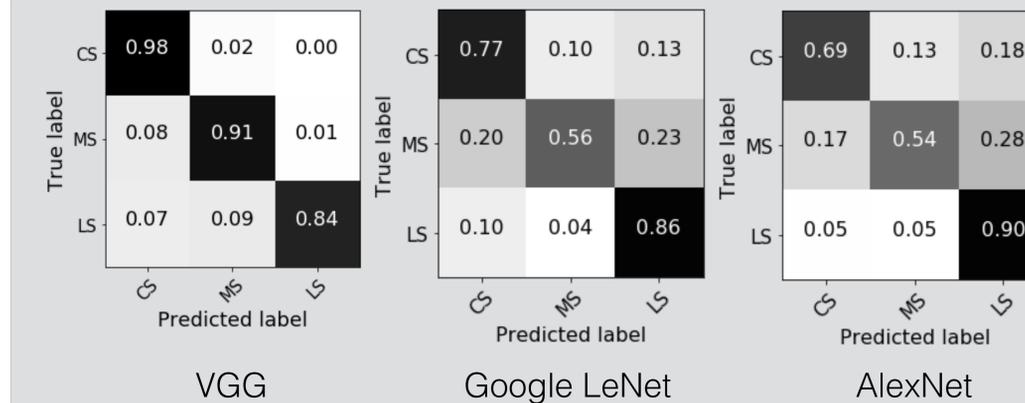


## CNN-Based Classification

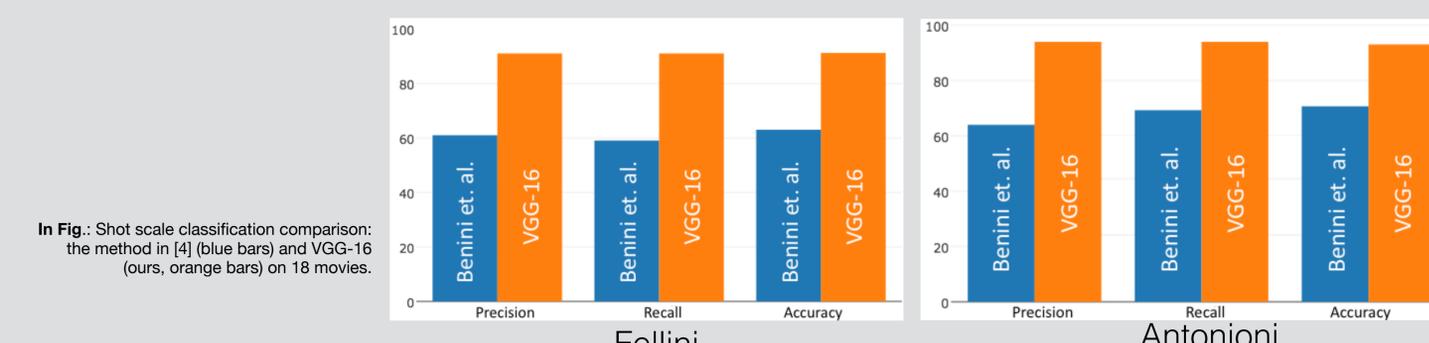


- Three tested deep learning (DL) architectures: AlexNet, GoogleLeNet, and VGG-16 [3].
- All CNNs perform better using precomputed weights from ImageNet. Two different networks (with the same architecture) are trained to deal with color and b/w movies, respectively.
- Exploitation of the temporal correlation of the shot scale: final moving window average of 3s.
- Training movies are 55, while the other 65 are used for testing. Movies in the two sets are chosen, whenever possible, by uniformly sampling director's productions, and trying to balance b&w and colour movies.

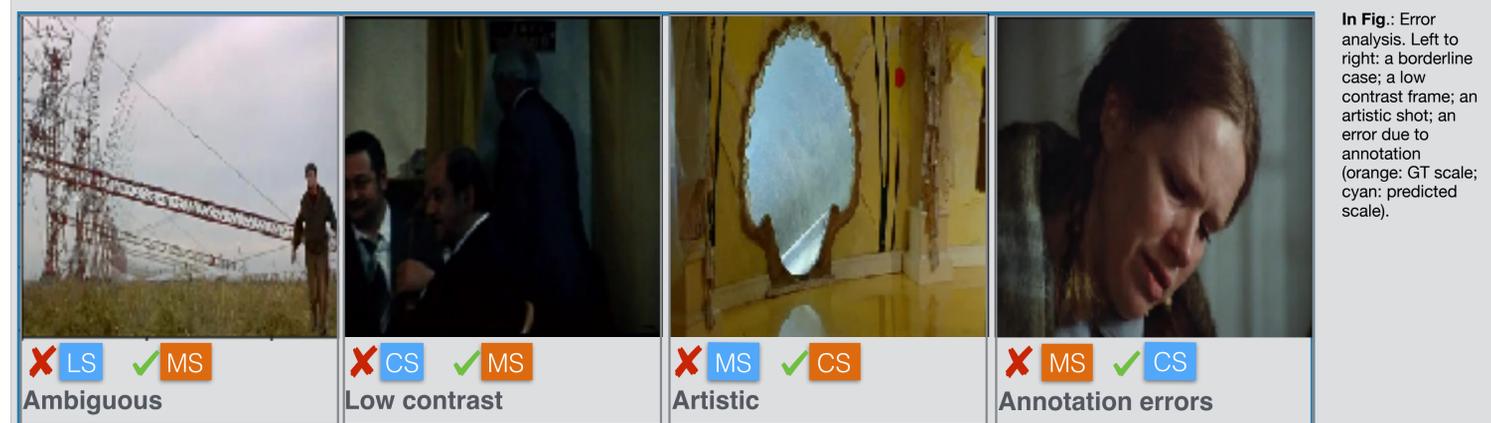
## Experiments



In Fig.: VGG-16 performs best with an overall precision of 94%, recall 94%, and accuracy 94%. For each CNN we test and compare four different configurations: A) training from scratch, B) loading weights from ImageNet and fine-tuning the last layer only, C) fine-tuning all fully-connected layers, and D) fine-tuning the whole network. Classification results for all other configurations using VGG-16 follow (in terms of accuracy): configuration A, 80%; configuration B, 89%; configuration C, 90%. The contribution of the post-processing smoothing increases the accuracy score of +0.5% on average.



In Fig.: Shot scale classification comparison: the method in [4] (blue bars) and VGG-16 (ours, orange bars) on 18 movies.



In Fig.: Error analysis. Left to right: a borderline case; a low contrast frame; an artistic shot; an error due to annotation (orange: GT scale; cyan: predicted scale).

## Conclusion

- In this work we propose a method for automatic classification of shot scale in three different classes (CU, MS, LS).
- The relevance of this study is motivated by the prominent aesthetic role of the shot scale and its emotional effects on viewers, abundantly described in cinema studies and psychology.
- Obtained classification accuracy on the three categories is superior to current state-of-the-art algorithms (>90%), which opens up new possibilities for interesting research applications at the crossroad between computer vision, cinema studies, and psychology.

## Main references

- [1] K. Balint, T. Klausch, and T. Polya, "Watching closely: Spatial distance influences theory of mind responding in film viewers," *Journal of Media Psychology: theories, methods, and applications*, pp. 1–10, 2016.
- [2] A. B. Kovacs, "Shot scale distribution: an authorial fingerprint or a cognitive pattern?" *Projections*, vol. 8, no. 2, 2014.
- [3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: <http://arxiv.org/abs/1409.1556>.
- [4] S. Benini, M. Svanera, N. Adami, R. Leonardi, and A. B. Kovacs, "Shot scale distribution in art films," *Multimedia Tools and Applications*, pp. 1–29, 2016.

