Utilizing a Convolutional Neural Network to Efficiently Label a Solar Flux Emergence Video Dataset

Subhamoy Chatterjee, Andrés Muñoz-Jaramillo, Derek Lamb
Southwest Research Institute, Boulder, CO, USA

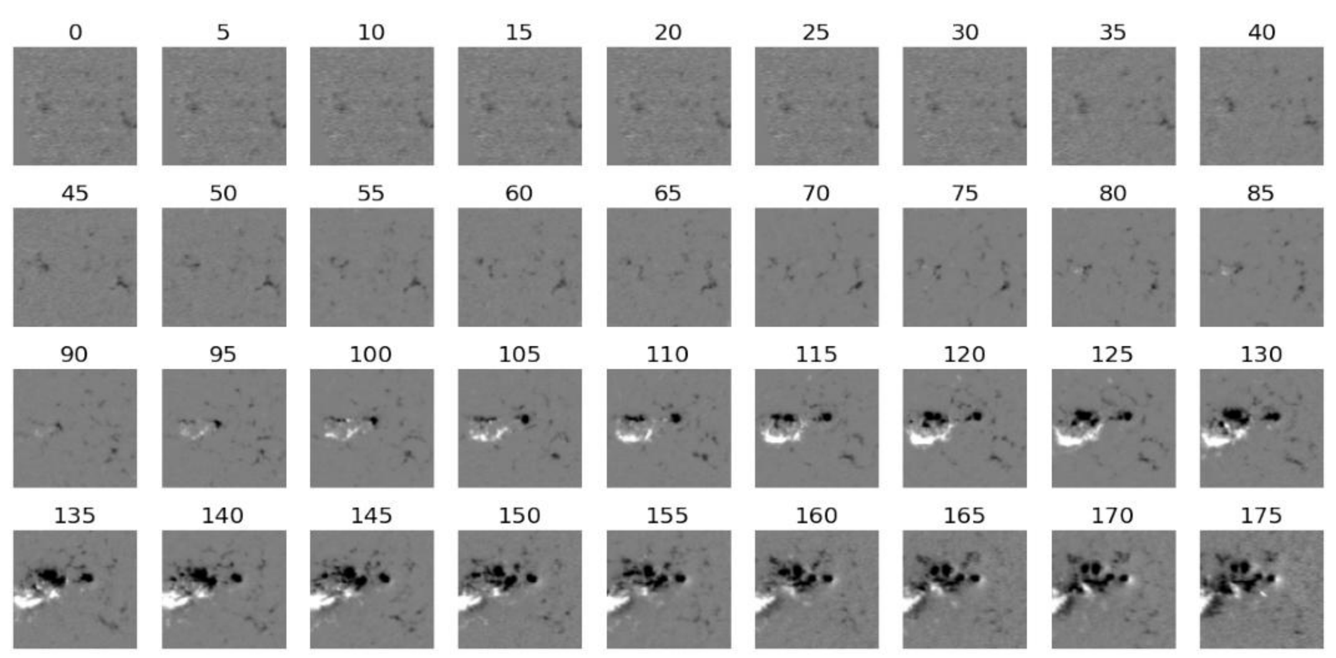
Funded by NASA Grants 80NSSC18K0671 & 80NSSC19M0165

INTRODUCTION & MOTIVATION

- Big-data problem in astronomy [1]
- Deep learning is showing promise in dealing with complex data - images, videos [2]
- Manual supervision demands time and consistency
- Iterative labelling approach such as 'active learning' [3] can save time
- We focus on a specific problem: 'flux emergence' depicted by the appearance of bipolar/complex magnetic regions on the solar surface and can have potential to drive solar events to modulate space weather and affect communication satellites [4]
- We start with a crudely labelled flux emergence data, train a deep learning model-ensemble and refine the labelling using the trained models as described through the rest of this poster

INITIAL DATASET

- SoHO/MDI [5] LOS magnetic patch videos
- Video size after interpolation 90x90x225 in Carrington longitude, latitude and time grid
- Crudely Labelled for active region 'emergence' and 'non-emergence' using BARD catalogue [6]
- Binary video classification problem



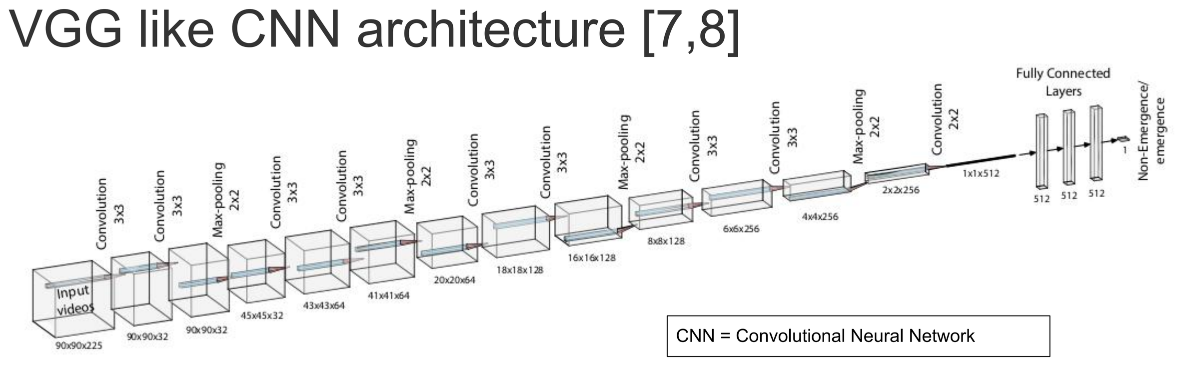
Abbreviations:

➔ *LOS = Line Of Sight*

➔ *BARD = Bipolar Active Region Detection*

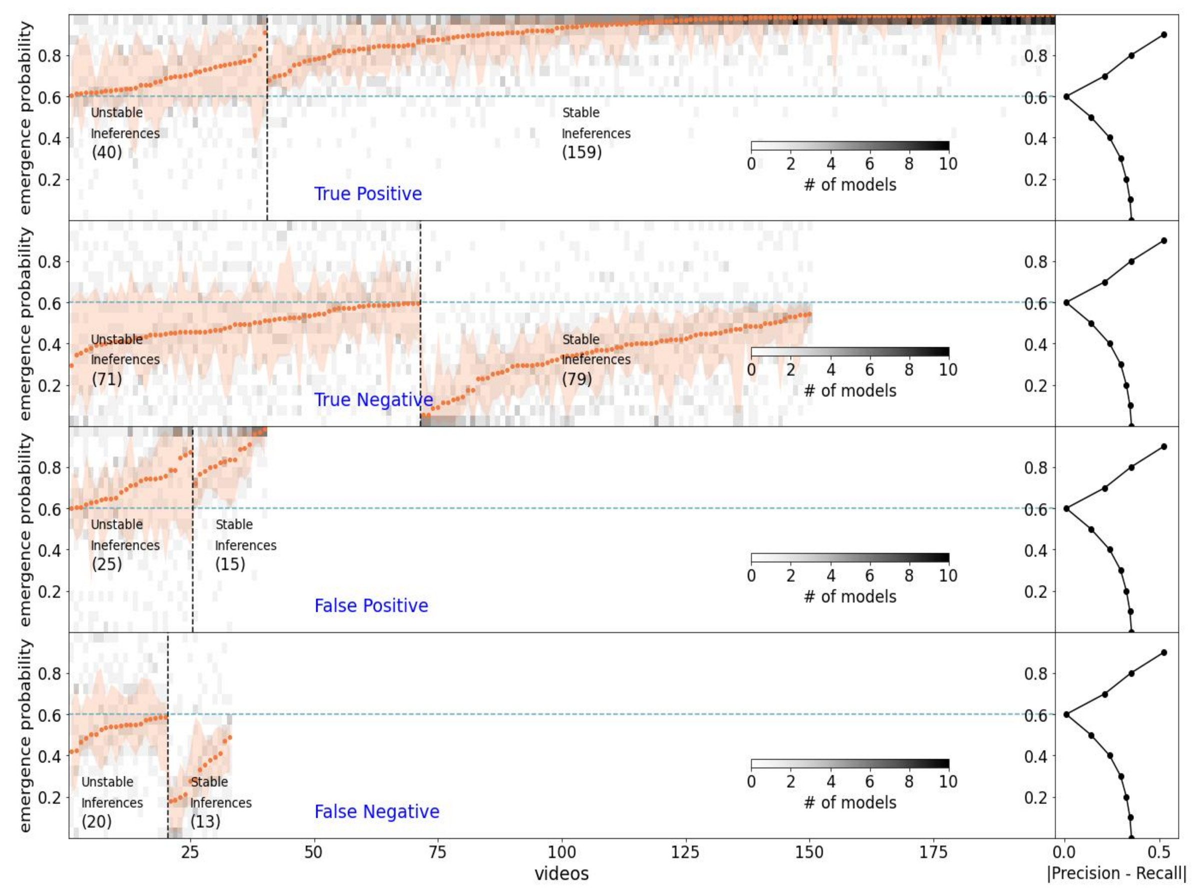
CNN MODEL & PERFORMANCE

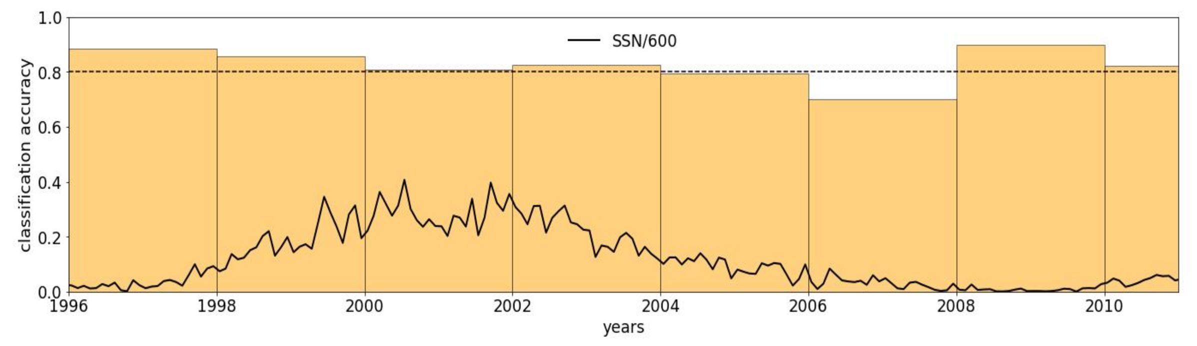
- VGG like CNN architecture [7,8]



- Training-validation combinations with random realisations: ensemble of 10 CNNs
- ~1600 videos for training and ~400 for validation

Months	Iterative Data Relabelling Performance Evaluation				
	model1	model2	model3	model4	model5
Jan	Training Set	Training Set	Training Set	Training Set	Validation
Feb	Training Set	Training Set	Training Set	Training Set	Validation
Mar	Training Set	Training Set	Training Set	Validation	Training Set
Apr	Training Set	Training Set	Training Set	Validation	Training Set
May	Training Set	Training Set	Validation	Training Set	Training Set
Jun	Training Set	Training Set	Validation	Training Set	Training Set
Jul	Training Set	Validation	Training Set	Training Set	Training Set
Aug	Training Set	Validation	Training Set	Training Set	Training Set
Sep	Validation	Training Set	Training Set	Training Set	Training Set
Oct	Validation	Training Set	Training Set	Training Set	Training Set
Nov	Test Set	Test Set	Test Set	Test Set	Test Set
Dec	Test Set	Test Set	Test Set	Test Set	Test Set

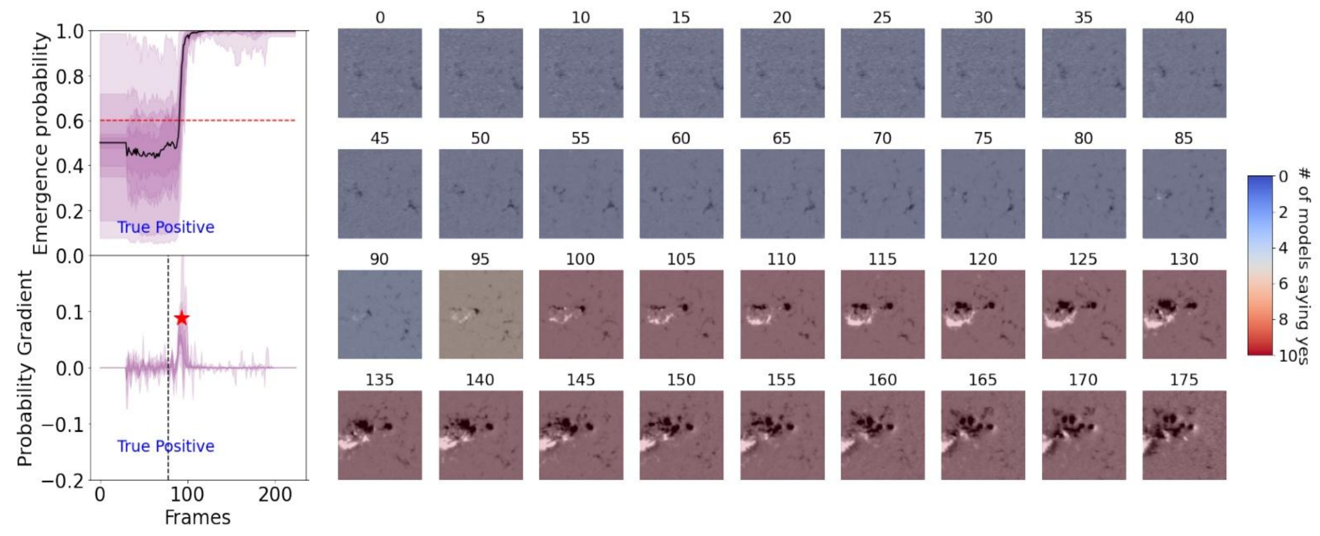




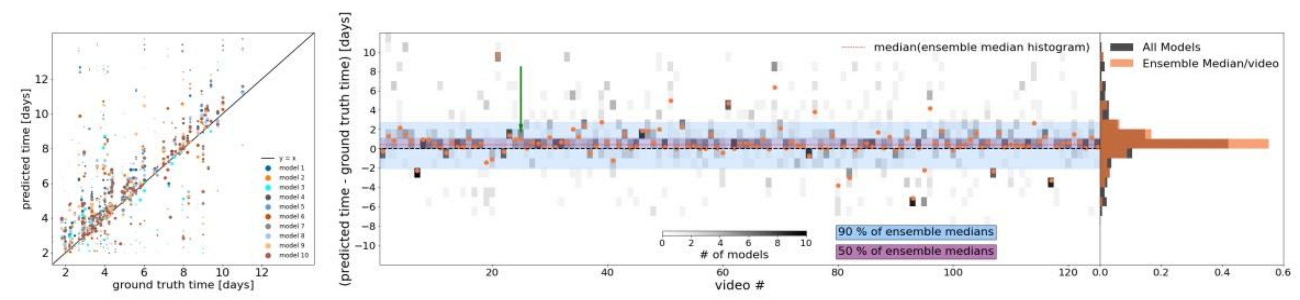
- We achieve 83% classification accuracy on the test set
- No systematic effect introduced by solar cycle phase on accuracy

MODEL REPURPOSING: FRAME LABELLING

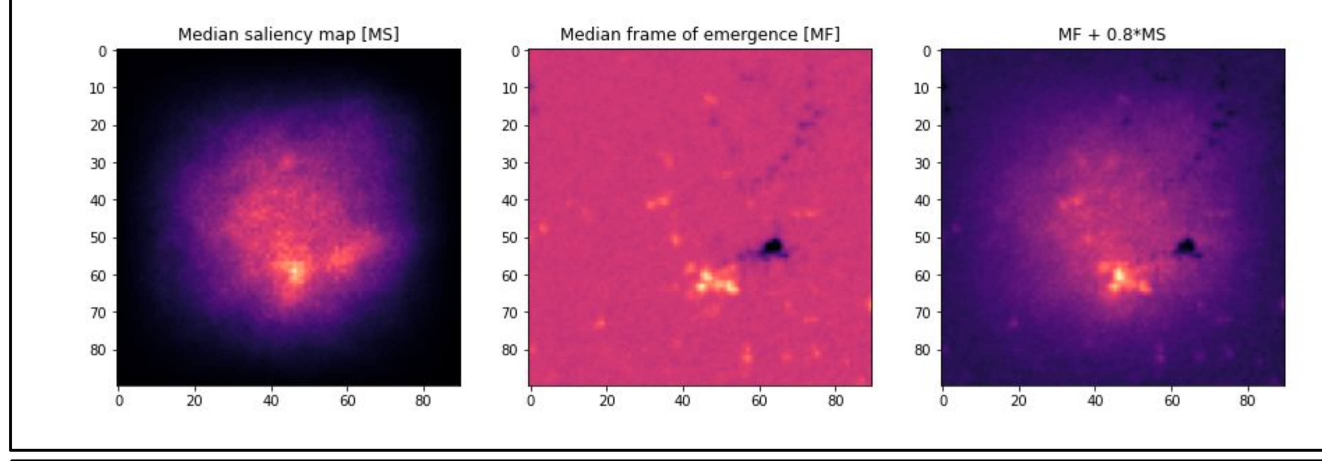
- We repurpose the model to pinpoint the time of emergence



- For this we truncate videos from the end until maximum change in inference

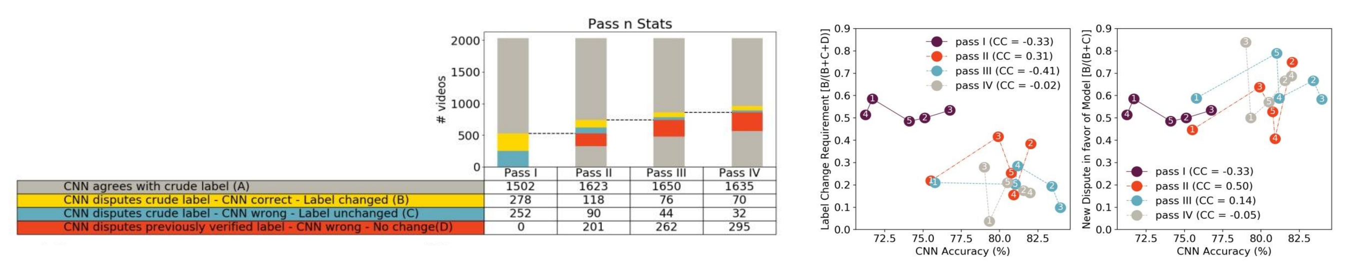


- Model identifies emergence times within [-1.6, 2.4] days from the ground truth times for 90% of the cases
- Saliency analysis to check if the model is giving importance to the right regions



ITERATIVE RELABELING

- Starting with crudely labelled data we train and evaluate the model-ensemble on respective validation blocks, manually verify model vs. crude label disagreements and repeat this process until convergence generating high quality labeled data



- After 4 iterations we correct 85% of mislabels by manually verifying only 50% of the data

CONCLUSION

- We classify MDI magnetic patch evolution into emergence or non-emergence
- Given the limited size of data points, we train a model-ensemble and reach a classification accuracy of 83%
- We also find that apart from video classification the model identify the time of emergence
- We show that the deep learning model can be used to refine the labelling of a crudely labelled dataset
- Usage of the model iteratively rules out the need of manually verifying the entire data

References

1. Zhang, Y. & Zhao, Y. Astronomy in the Big Data Era.Data Sci. J.14, 11 (2015). DOI 10.5334/dsj-2015-011.
2. Krizhevsky, A., Sutskever, I. & Hinton, G. E. In Pereira, F.,Burges, C. J. C., Bottou, L. & Weinberger, K. Q. (eds.)Advances in Neural Information Processing Systems 25, 1097–1105(Curran Associates, Inc., 2012).
3. Settles, Burr.. Computer Sciences Technical Report 1648. University of Wisconsin–Madison (2010).
4. Dubey, G., van der Holst, B. & Poedts, S..A & A459,927–934 (2006). DOI 10.1051/0004-6361:20054719
5. Scherrer, P. H.et al.T.162, 129–188 (1995).DOI 10.1007/BF00733429.
6. Mu˜noz-Jaramillo, A. et al.. IEEE International Conference on Big Data (Big Data)(2016)
7. LeCun, Y., Haffner, P., Bottou, L. & Bengio, Y., 319–345 (SpringerBerlin Heidelberg, Berlin, Heidelberg, 1999). DOI 10.1007/3-540-46805-6-19
8. K. Simonyan, A. Zisserman.. International Conference on Learning Representations (2015).