Prediction of Off-Fault Deformation from Experimental Strike-slip Fault Structures using the Convolutional Neural Networks

L. Chaipornkaew[1], H. Elston[2], M. Cooke[2], T. Mukerji[1], S.A. Graham[1]

[1]Stanford University

[2]University of Massachusetts Amherst

Corresponding author: Laainam Chaipornkaew (bestc@stanford.edu)

Key Points:

- Evolving experimental strike-slip fault maps are suitable for machine learning. The labels are calculated via image processing.
- Proposed convolutional neural networks (CNNs) can predict off-fault deformation directly from experimental fault trace maps.
- Trained CNN performs with 91% accuracy on unseen experimental faults and show promise in predicting kinematic efficiency of crustal faults.

Abstract

Crustal deformation occurs both as localized slip along faults and distributed deformation off faults; however, we have few robust estimates of off-fault deformation. Scaled physical experiments simulate crustal strike-slip faulting and allow direct measurement of fault slip to regional deformation, quantified as Kinematic Efficiency (KE). We offer an approach for KE prediction using a 2D Convolutional Neural Network (CNN) trained directly on images of fault maps produced by physical experiments. A suite of experiments with different loading rate and basal boundary conditions, contribute over 13,000 fault maps throughout strike-slip fault evolution. Strain maps allow us to directly calculate KE and its uncertainty, utilized in the loss function and performance metric. The trained CNN achieves 91% accuracy in KE prediction of an unseen dataset. The application of the CNN trained on scaled experiments to crustal fault maps provides estimates of off-fault deformation that overlap available geologic estimates.

Plain Language Summary

Where the earth deforms at the boundaries between tectonic plates, some of the deformation is taken up as localized slip along fault surfaces and some of the deformation is distributed around the fault. This distributed deformation is very hard to measure in the Earth's crust. To get around this challenge, we create faults in the laboratory and use the direct measurements of the distributed deformation off of faults to train a machine learning model. The trained model performs well at predicting distributed off-fault deformation from the fault geometry.

1 Introduction

Despite abundant documentation of distributed deformation, such as folding and fracturing, within bedrock and soil away from primary fault surfaces, we have few estimates of the portion of regional deformation that is permanently accommodated off of faults. Within regions of strike-slip faulting, geologic investigations of structures that record cumulative deformation suggest that ~10-30% of the regional strike-slip strain may occur off of individual strike-slip faults (e.g., (Goren et al., 2015; Gray et al., 2018; Shelef & Oskin, 2010; Titus et al., 2011). Comparisons of coseismic surface slip to geodetic estimates show greater strike slip rate discrepancy for immature faults with < 25 km cumulative slip than more mature faults (Dolan & Haravitch, 2014). Attributing the discrepancy in fault slip rate to off-fault deformation, Dolan & Haravitch (2014) find that immature faults have 40-50% off-fault deformation while mature faults (total slip > 25 km) have 10-15% off-fault deformation. Fault geometry, which is an outcome of fault maturity, exerts primary control on the portion of regional deformation partitioned as fault slip (C. Milliner et al., 2016). Strike-slip faults with smoother traces can more efficiently accommodate slip with lesser off-fault deformation than faults with rough/complex traces in the slip direction (e.g., Chester & Chester, 2000; Fang & Dunham, 2013; Newman & Griffith, 2014; Saucier et al., 1992).

The hypothesis that smoother faults produce greater slip is strongly supported by scaled physical experiments of strike-slip fault evolution that directly document that as faults mature and become smoother, the % of fault slip quantified as kinematic efficiency (KE = 1 - % off fault deformation) increases. KE describes the ratio of incremental strike-slip accommodated along the fault to the total incremental displacement (Hatem et al., 2017). Arrays of echelon faults that develop early in strike slip fault evolution have KE of 50-60% and this increases to 80-90% along the through-going mature strike-slip fault (Hatem et al., 2017). The direct and complete observations available from experiments can inform our interpretations of crustal faulting in ways that field data, which generally only reveal cumulative deformation and parts of the structure, cannot (e.g., Reber et al., 2020). The power of scaled experiments derives from using carefully selected analog materials that allow fault evolutionary processes that occur over very large space and time scales to be simulated in a few hours within the laboratory.

What controls the kinematic efficiency of strike-slip faults? Studies of both the coseismic deformation fields and experiments show correlations of amount of fault slip with both fault zone width (Hatem et al., 2017; C. W. Milliner et al., 2015) as well as strike-slip fault trace roughness/complexity (Hatem et al., 2017; C. Milliner et al., 2016). While fractal dimension can quantify the roughness of continuous fault traces (e.g., Brown, 1987) this metric cannot reliably capture the roughness of segmented faults where connectivity controls fault slip. Because any single metric will overlook aspects that may relate strike-slip fault architecture and KE, we need an alternative approach.

In this study, we harness a machine learning algorithm on an experimental time series of fault maps to estimate kinematic efficiency. Experiments that are scaled to simulate crustal strike-slip fault development allow direct detailed observation of the evolution of both active fault network and KE under a range of loading rates and boundary conditions. Machine learning has been used to predict the timing and size of lab quakes produced in rock (e.g., Rouet-Leduc et al., 2017) and scaled analog materials (Corbi et al., 2019). We use Convolutional Neural Networks (CNNs), which have proved successful for a wide range of computer vision tasks (e.g.,LeCun et al., 2010) because they can relate relevant parameters in higher dimensions to specific prediction tasks. With experimental strike-slip fault dataset, our CNNs associate the complexity of the active fault network with the degree of off-fault deformation.

2 Data and Methodology

2.1 Experimental fault data

We record the changing complexity of strike-slip faults during their evolution within experiments under various boundary conditions that represent different crustal loading. The table-top deformation apparatus consists of a split box filled with wet kaolin clay (Figure 1a). The benefits of wet kaolin over other crustal analog materials are that 1) wet kaolin creates very clear faults that can be tracked (e.g., Eisenstadt & Sims, 2005; Henza et al., 2010; Oertel, 1965; Tchalenko, 1970), 2) the low but non-zero cohesion of wet kaolin ensures that faults are long-lived; (compared to dry granular materials (Cooke et al., 2013; Eisenstadt & Sims, 2005; Withjack et al., 2007) and 3) the viscoelastic behavior of wet kaolin can simulate off fault relaxation of stresses with the crust (Cooke & van der Elst, 2012). Many studies have used kaolin to simulate evolution of strike-slip fault systems (Cooke et al., 2013; Hatem et al., 2015, 2017; Tchalenko, 1970).

The scaling of the wet kaolin and details of the experimental set up and analysis are described in the supplement to this paper. We simulated two basal shear conditions: distributed shear via a 2.5 cm wide elastic band secured to the basal plates or localized shear by juxtaposing the two basal plates. Computer-controlled stepper motors displaced one half of the box at a prescribed rate (0.25, 0.5, 1.0, or 1.5 mm/min) parallel to the basal plate discontinuity to initiate dextral strike-slip faulting within the overlying wet kaolin. We conducted experiments at least twice for each condition for a total of 16 experiments.

The experimental active fault maps capture the evolution of strike-slip faults from echelon crack initiation to through-going slip surface. Because digital image correlation techniques provide us with precise knowledge of the incremental horizontal displacement fields, we can directly calculate localized slip along the faults and kinematic efficiency (KE labels) across any portion of the fault. Additionally, we record the associated standard deviation (SD) of KE across the region input to the CNN. Both the KE labels and SDs are utilized in the optimization of the CNN (Eq. 1) and assessment of the CNN performance (Eq. 2)

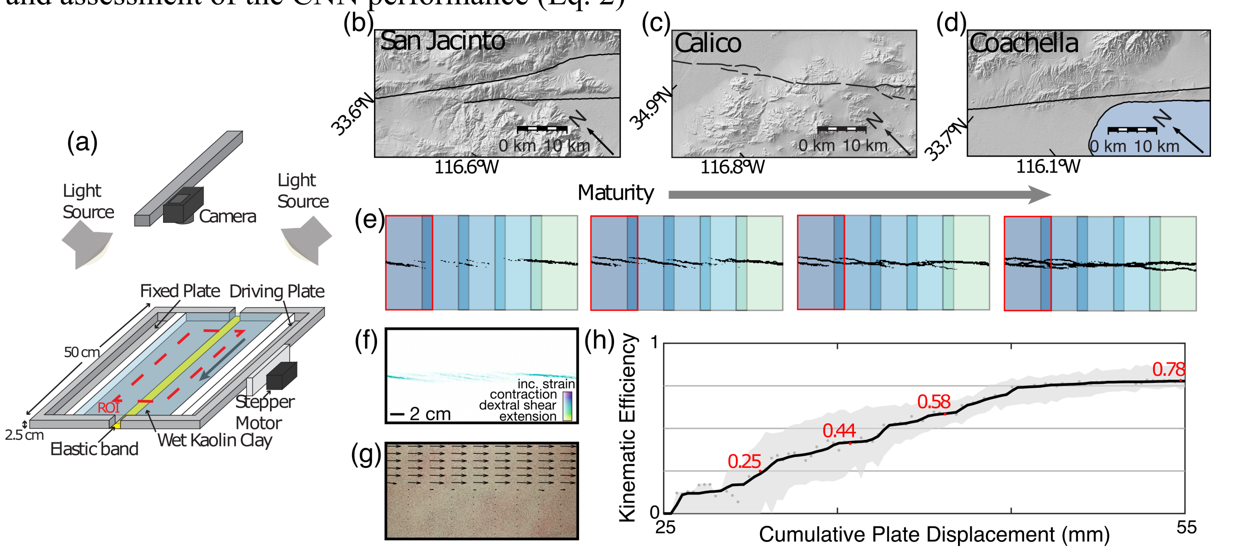

Figure 1: (a) Schematic of the distributed basal shear experiment loaded in strike-slip. (b) incremental displacement vectors at a snapshot of the fault evolution. (c) The shear strain map derived from (b). (d) Example experiment fault maps (1.5 mm/min distributed basal shear). Color shading delineates individual windows and their overlap. (e) KE for the experiment increases with strike-slip fault maturity. The grey band indicates the range of KE within individual windows along the experimental fault. The red numbers report KE for specific example windows outlined in red in (d). (f – h) Strike slip fault traces from Southern California show how complexity changes with increasing fault maturity along the (f) San Jacinto fault (map center at 33.45°N 116.45°W), (g) Calico fault (map center at 34.65°N 166.6°W), and (h) Coachella segment of the San Andreas fault near Mecca Hills (map center at 33.58°N 116.95°W).

2.2 Data Processing

We subsample each fault map into five 128x64 pixel windows with ~ 20% overlap (Fig. 1e). This window size captures multiple echelon strands during early stages of fault development while allowing larger numbers of unique datasets that are essential for training and testing of the model. We divide the complete dataset into three statistically equivalent subsets for training (64%), evaluation (24%), and test (12%) purposes. Since each experiment set-up is repeated, all 5 windows of the first experiments of each set-up are included in the training dataset while the 5 windows within the repeated runs are randomly allocated for training (⅖), evaluation (⅖), and testing (⅕). With this approach, each dataset contains all ranges of boundary conditions. We use the unseen test dataset to assess the trained CNN performance.

After filtering out early maps prior to the initiation of active faults, the dataset consists of approximately 13,400 sliced windows. In order to generate a sufficiently large dataset, we apply common geometric augmentation techniques (e.g., Shorten & Khoshgoftaar, 2019). Trivial transformation does not contribute to significant diversification of the dataset while over-exaggerated transformation distorts the fault traces so much that the CNN cannot associate fault maps to kinematic efficiency. We found that zooming (+/- 5%), shifting (+/- 10%), and flipping (binary) each window are the most effective ranges of transformation for the dimension and scaling of our dataset. Augmented fault locations, sizes, and slip sense better represent broad varieties of potential crustal active fault maps. See supplementary to this paper for further details of the data processing.

2.3 CNN Methodology

Convolutional Neural Networks (CNNs) trained using experimental strike-slip fault maps can provide a useful way to describe the complex and non-linear relationship between active fault trace complexity and kinematic efficiency. Learning directly from fault maps eliminates the need to prescribe exact equations to describe complex failure behaviors. The proposed CNNs learn how active fault traces relate to KE by minimizing a custom loss function L based on a normalized mean square error as shown in Eq. 1

$$L = \frac{MSE}{SD^2 + \epsilon}\ ; MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\ , \epsilon = 1e^{-5} \tag{1}$$

The mean square error (MSE) is the squared difference of the estimated values (KE prediction, $\hat{y}_i$) and the truth (KE label, y_i). A small value of ensures a non-zero divisor. Our custom loss function scales MSE with the squared standard deviation of KE (SD), allowing the model to learn more precisely where we have the most confidence while relaxing the learning conditions where uncertainties are high.

We train the models with Adam Optimizer (Kingma & Ba, 2017), a modified stochastic gradient descent with adaptive learning rate. Our CNN network (Fig 2a) has stacked convolutional layers that have appropriate kernel sizes and dilation parameters to detect features both locally (i.e. stepovers between faults) and globally (fault connectedness). We then apply batch normalization, which stabilizes learning with a modest regularization effect that improves the performance (Ioffe & Szegedy, 2015). We chose the rectified linear (ReLU), as our non-linear activation function, to enable the network to approximate functions and preserve properties of each feature map (Kulathunga et al., 2021; Nair & Hinton, 2010). Subsequently, we use max-pooling to highlight the most representative features and then downsample the feature maps so that they would not be sensitive to the faults' location within the input maps. Pooled layers also reduce the number of parameters, and aid computing efficiency before being flattened into a vector in fully-connected layers. We add a dropout regularizer (Srivastava et al., 2014) to improve generalization and prevent overfitting in predicting KE.

We iterate over a range of network depth and select the most efficient network (Fig. 2a) that performs well. We assess the performance of our CNN networks by considering the prediction as correct if the absolute difference of the predicted KE and the true KE label fit within two standard deviations of the label (Eq. 2).

$$Correct\ if\ \ |y_i - \hat{y}_i|\ <\ 2SD \tag{2}$$

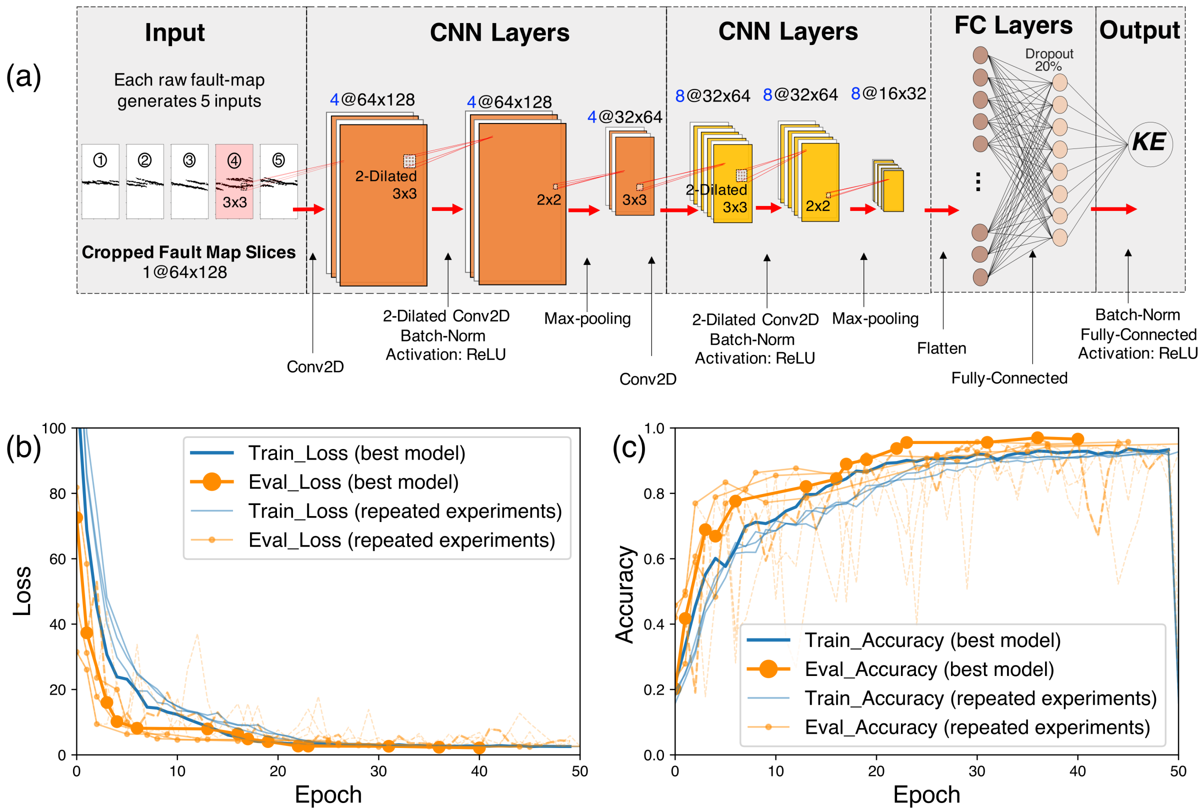

Figure 2: (a) Convolutional neural network architecture designed to optimize for efficiency and performance. (b-c) Loss and accuracy as we fit the models over 50 training epochs and evaluate the performance against the loss and accuracy at the end of each epoch. The solid lines represent the best performing model. Dotted lines represent the results of comparable fitting using the same set of hyperparameters showing repeatability of all models, achieving consistently high performances.

3 Results

3.1 CNN Prediction on Experimental Faults

To ensure that the trained CNN can generalize to unseen data, we use the minimum loss (Eq. 1) of the evaluation dataset to guide tuning of the hyperparameters. The best model, and all repeated training runs illustrate a good fit, and the CNN model stops improving after approximately 50 training epochs, where we impose an early stopping of the training process(Fig. 2b). Additionally, we confirm the repeatability of the models' performance by reproducing mini-batch accuracy over 90% (Table S3) on all training sessions using the same set of hyperparameters (Table S2) while varying the randomized initialization.

Applying the selected CNN's model (best model) for prediction tasks, we reach high performance of 96.7% and 96.1% accuracy (Eq.2) in training and evaluation datasets respectively (Fig 2C). Similarly, prediction on an unseen test dataset yields satisfactory performance of 90.9% accuracy. These correct predictions for the majority of the dataset extensively represent experiments with the full range of applied loading rates, basal boundary conditions and stages of fault evolution. On the other hand, the clusters of outliers from more matured faults seem to correlate to individual experiments within a specific KE range (Fig. 3).

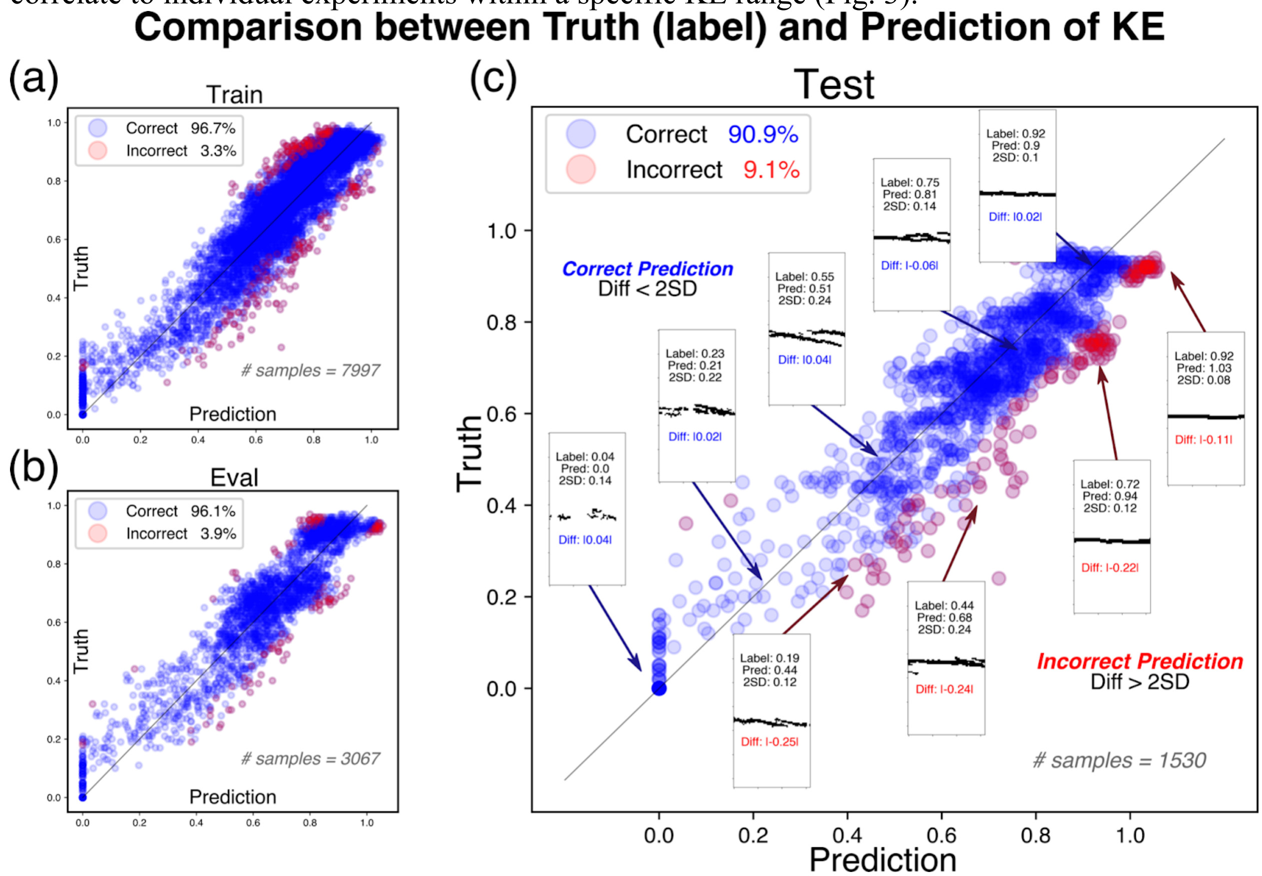

Figure 3: Prediction performance on training, evaluation, and test datasets using the selected model that generates the lowest evaluation loss during training for the prediction task. (a)-(b) Trained CNN can predict KE of both training and evaluation dataset with an accuracy exceeding 96%. (c) Trained CNN can predict KE of an unseen test dataset with 91% accuracy. On the upper left, we display selected fault maps from all KE ranges. These examples represent datasets that can be accurately predicted by CNN. On the lower right, we display examples of a few outliers that cannot be predicted correctly by the trained CNN.

3.2 CNN Application on Crustal Fault Maps

Because there are very few geologic estimates of off-fault deformation in the crust, we strictly train the CNN model using only experimental faults. But since these laboratory simulated faults are scaled to crusts, the trained CNN has potential to predict KE of crustal faults. Here, we compare the off-fault deformation estimates from three geologic studies to CNN predicted KE that use the active fault maps of those studies. These studies use evidence of off-fault deformation accumulated across different time spans: coseismic deformation associated with 2016 M7 Kumamoto earthquake along the Futagawa fault (Scott et al., 2019), Holocene deformation of drainages adjacent to the San Andreas fault at Mecca Hills (Gray et al., 2018) and bending of Mesozoic faults and dikes adjacent to the northern Calico fault in the Eastern California Shear Zone (Shelef & Oskin, 2010). To prevent misinterpreting a mapped fault trace thickness, which varies with map scale and may be correlated to fault maturity, we have the CNN predict KE for a range of fault trace thickness (see supplement). The CNN predicts KEs from the mapped fault traces that overlap all of the geologic estimates (Figure 4). The CNN predicts the greatest off-fault deformation (lowest KE) for the Calico fault because its trace is the most segmented; the CNN predicts the least off-fault deformation for the San Andreas, which is the most straight and continuous. The segmentation of the Calico fault trace produces along-strike variability of its trace that contributes to the wide range of CNN predicted KEs.

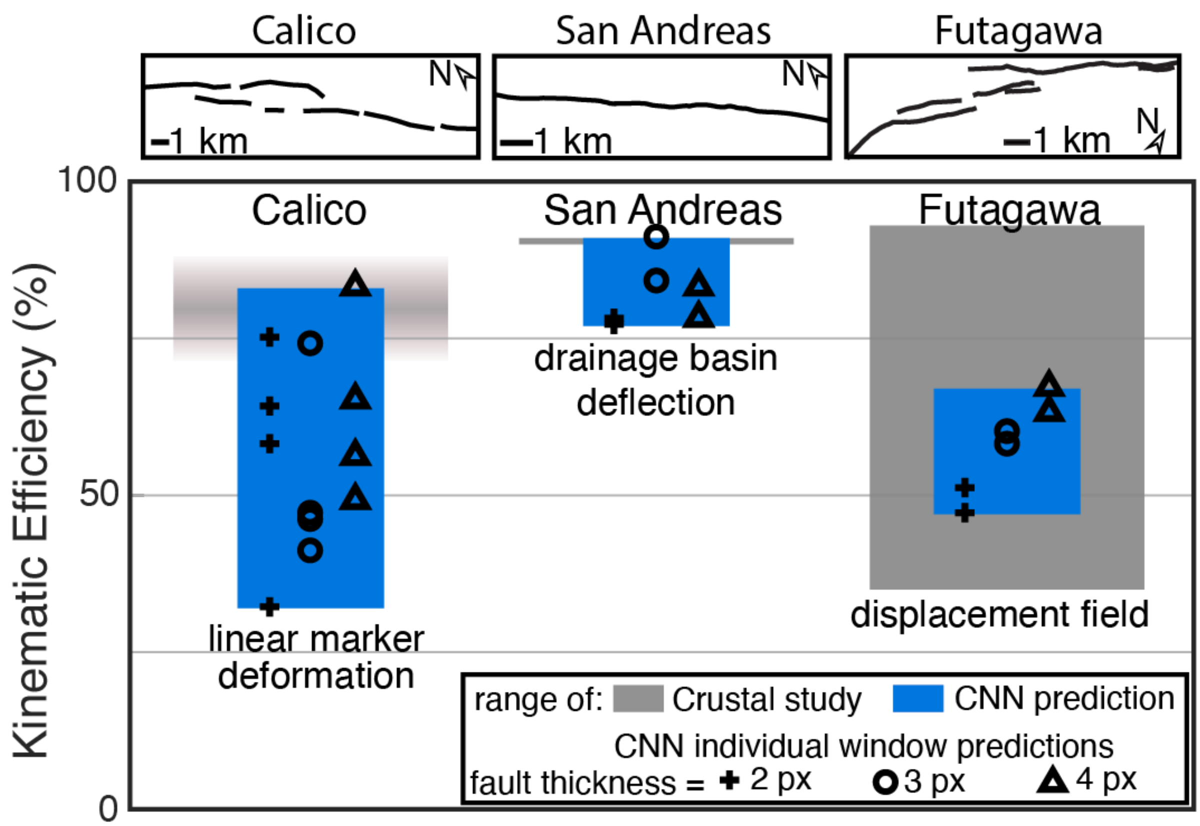

Figure 4. Crustal fault trace maps and the kinematic efficiency ranges from geologic studies (grey rectangles with uncertainty represented by shading) and the CNN predictions (blue rectangles). The different symbols indicate the CNN's prediction for three different representations of fault thicknesses, which might vary with map scale. For example, 2 pixels within the dataset used to train the CNN scales to 200 m in the crust.

Interestingly, the range of off-fault deformation estimates vary considerably between the geologic studies due to the different methods employed. The range of off-fault estimates for coseismic deformation along the Futagawa fault may be larger than those of Gray et al. (2018) and Shelef & Oskin (2010) because the digital image correlation methods used by Scott et al. (2019) allow for more complete detection of off-fault deformation than available from either geomorphic or geologic records. Having the complete displacement field captures spatial variations in off-fault deformation whose fingerprint might be averaged within the landscape and within the permanent deformation of rocks adjacent to the fault. The results here show the potential for the CNN to estimate crustal fault KE from fault traces alone.

4 Discussion

The CNN successfully trained here provides accurate prediction of KE and corresponding off-fault deformation of experimental faults. A combination of the optimal hyperparameters, customized loss function, and efficient network architecture contribute to the satisfactory performance of the trained model to predict KE in experimental strike-slip faults. Performance and learnability of the CNNs are impacted by the scale of geometric augmentation, especially the stretching factor. This indicates a sensitivity of scaling and thickness of crustal fault maps that the CNN is able to predict.

The width of the map window used for the CNN development can impacts both the KE label and the CNN predicted KE of segmented faults. For example, the segmentation of faults may not be fully captured within the windows that we used in this study. If the window width does not span adjacent fault segments but only captures a single segment, the calculated and predicted KE will be higher than along the segmented fault. Wider windows can sample step overs more reliably. Within the strike-slip experiments, the kinematic efficiency within a single 6.4 cm window has standard deviation up to 20%, demonstrating that KE varies significantly over a relatively short distance along faults.

While we expect that more mature faults are more localized than immature faults (e.g., Tchalenko, 1970), and strike-slip experiments show decreasing shear zone with fault maturity (Hatem et al., 2017), the CNN predicted KE for the crustal fault maps did not vary systematically with fault trace thickness (Fig. 4). Furthermore, the kaolin used here produces highly localized faults whereas other crustal analog materials, such as sand, produce wider zones of faulting and may produce different degrees of off-fault deformation (Reber et al., 2020). The degree of off-fault deformation within viscoelastic materials, such as the kaolin used here, depends on applied loading rate. Because we trained the CNN on experiments with a range of loading rates, the results here incorporate some degree of off-fault deformation variability. Because the active morphology of crustal strike-slip fault may owe to processes not captured in the wet kaolin experiments of this study, the CNN trained here may not accurately predict off-fault deformation of all strike-slip faults. Retraining the CNN on additional data sets may produce more robust predictive tools for crustal faults.

5 Conclusion

Because seismic hazard analyses benefit from estimates of off-fault deformation, we need reliable ways to quantify the portion of strain that is accommodated off of faults. Here, we offer an alternative approach for Kinematic Efficiency (KE) prediction in strike-slip using a 2D Convolutional Neural Network, that is trained directly on images of fault maps produced by fault experiments scaled to simulate crustal strike-slip faults. Our dataset captures the whole evolution of strike-slip faults and allows precise calculation of off-fault deformation (1-KE). We use a custom loss function and custom accuracy, which fully utilize both the KE labels and their standard deviation. We tune the set of hyperparameters to optimize our CNN training. The final CNN model has the ability to predict on an unseen test dataset with 91% accuracy. Lastly, the match of the CNN to crustal fault maps with off-fault deformation estimates shows the potential for applying experimentally trained CNNs to crustal faults.

Data Availability Statement

We have submitted the experimental PIV experiment data and strain map animations of all experiments to the EPOS analog modeling repository at GFZ (Cooke et al, 2021). Dataset to this manuscript is published to EPOS data repository at https://doi.org/10.5880/GFZ.fidgeo.2021.029. Because the EPOS DOI link provided is not yet activated, the authors temporarily upload our dataset as Supporting Information for review purposes. The codes and selected models used in this paper are available via GitHub Repository, deposited at https://doi.org/10.5281/zenodo.5155156

Acknowledgments

The authors thank Stanford's Basin and Petroleum System Modeling Group for technical advice and funding of graduate study to L. Chaipornkaew. This experimental material is based on work partially supported by the US Geological Survey grant G21AP10398 to M. Cooke and GSA Graduate Student Research Grant to H. Elston.

References

Brown, S. R. (1987). A note on the description of surface roughness using fractal dimension. *Geophysical Research Letters*, *14*(11), 1095–1098.

Chester, F. M., & Chester, J. S. (2000). Stress and deformation along wavy frictional faults. *Journal of Geophysical Research: Solid Earth*, *105*(B10), 23421–23430.

Cooke, M. L., & van der Elst, N. J. (2012). Rheologic testing of wet kaolin reveals frictional and bi-viscous behavior typical of crustal materials. *Geophysical Research Letters*, *39*(1).

Cooke, M. L., Schottenfeld, M. T., & Buchanan, S. W. (2013). Evolution of fault efficiency at restraining bends within wet kaolin analog experiments. *Journal of Structural Geology*, *51*, 180–192.

Corbi, F., Sandri, L., Bedford, J., Funiciello, F., Brizzi, S., Rosenau, M., & Lallemand, S. (2019). Machine learning can predict the timing and size of analog earthquakes. *Geophysical Research Letters*, *46*(3), 1303–1311.

Dolan, J. F., & Haravitch, B. D. (2014). How well do surface slip measurements track slip at depth in large strike-slip earthquakes? The importance of fault structural maturity in controlling on-fault slip versus off-fault surface deformation. *Earth and Planetary Science Letters*, *388*, 38–47.

Eisenstadt, G., & Sims, D. (2005). Evaluating sand and clay models: do rheological differences matter? *Journal of Structural Geology*, *27*(8), 1399–1412.

Fang, Z., & Dunham, E. M. (2013). Additional shear resistance from fault roughness and stress levels on geometrically complex faults. *Journal of Geophysical Research: Solid Earth*, *118*(7), 3642–3654.

Goren, L., Castelltort, S., & Klinger, Y. (2015). Modes and rates of horizontal deformation from rotated river basins: Application to the Dead Sea fault system in Lebanon. *Geology*, *43*(9), 843–846.

Gray, H. J., Shobe, C. M., Hobley, D. E., Tucker, G. E., Duvall, A. R., Harbert, S. A., & Owen, L. A. (2018). Off-fault deformation rate along the southern San Andreas fault at Mecca Hills, southern California, inferred from landscape modeling of curved drainages. *Geology*, *46*(1), 59–62.

Hatem, A. E., Cooke, M. L., & Madden, E. H. (2015). Evolving efficiency of restraining bends within wet kaolin analog experiments. *Journal of Geophysical Research: Solid Earth*, *120*(3), 1975–1992.

Hatem, A. E., Cooke, M. L., & Toeneboehn, K. (2017). Strain localization and evolving kinematic efficiency of initiating strike-slip faults within wet kaolin experiments. *Journal of Structural Geology*. https://doi.org/10.1016/j.jsg.2017.06.011

Henza, A. A., Withjack, M. O., & Schlische, R. W. (2010). Normal-fault development during two phases of non-coaxial extension: An experimental study. *Journal of Structural Geology*, *32*(11), 1656–1667.

Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *ArXiv:1502.03167 [Cs]*. Retrieved from http://arxiv.org/abs/1502.03167

Kingma, D. P., & Ba, J. (2017). Adam: A Method for Stochastic Optimization. *ArXiv:1412.6980 [Cs]*. Retrieved from http://arxiv.org/abs/1412.6980

Kulathunga, N., Ranasinghe, N. R., Vrinceanu, D., Kinsman, Z., Huang, L., & Wang, Y. (2021). Effects of Nonlinearity and Network Architecture on the Performance of Supervised Neural Networks. *Algorithms*, *14*(2), 51. https://doi.org/10.3390/a14020051

LeCun, Y., Kavukcuoglu, K., & Farabet, C. (2010). Convolutional networks and applications in vision. In *Proceedings of 2010 IEEE international symposium on circuits and systems* (pp. 253–256). IEEE.

Milliner, C., Sammis, C., Allam, A., Dolan, J., Hollingsworth, J., Leprince, S., & Ayoub, F. (2016). Resolving fine-scale heterogeneity of co-seismic slip and the relation to fault structure. *Scientific Reports*, *6*(1), 1–9.

Milliner, C. W., Dolan, J. F., Hollingsworth, J., Leprince, S., Ayoub, F., & Sammis, C. G. (2015). Quantifying near-field and off-fault deformation patterns of the 1992 Mw 7.3 L anders earthquake. *Geochemistry, Geophysics, Geosystems*, *16*(5), 1577–1598.

Nair, V., & Hinton, G. E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning* (pp. 807–814). Madison, WI, USA: Omnipress.

Newman, P. J., & Griffith, W. A. (2014). The work budget of rough faults. *Tectonophysics*, *636*, 100–110.

Oertel, G. (1965). The mechanism of faulting in clay experiments. *Tectonophysics*, *2*(5), 343–393.

Reber, J. E., Cooke, M. L., & Dooley, T. P. (2020). What model material to use? A Review on rock analogs for structural geology and tectonics. *Earth-Science Reviews*, *202*, 103107.

Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., & Johnson, P. A. (2017). Machine learning predicts laboratory earthquakes. *Geophysical Research Letters*, *44*(18), 9276–9282.

Saucier, F., Humphreys, E., & Weldon, R. (1992). Stress near geometrically complex strike-slip faults: Application to the San Andreas fault at Cajon Pass, southern California. *Journal of Geophysical Research: Solid Earth*, *97*(B4), 5081–5094.

Scott, C., Champenois, J., Klinger, Y., Nissen, E., Maruyama, T., Chiba, T., & Arrowsmith, R. (2019). The 2016 M7 Kumamoto, Japan, earthquake slip field derived from a joint inversion of differential lidar topography, optical correlation, and InSAR surface displacements. *Geophysical Research Letters*, *46*(12), 6341–6351.

Shelef, E., & Oskin, M. (2010). Deformation processes adjacent to active faults: Examples from eastern California. *Journal of Geophysical Research: Solid Earth*, *115*(B5).

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, *6*(1), 60. https://doi.org/10.1186/s40537-019-0197-0

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, *15*(56), 1929–1958.

Tchalenko, J. (1970). Similarities between shear zones of different magnitudes. *Geological Society of America Bulletin*, *81*(6), 1625–1640.

Titus, S. J., Crump, S., McGuire, Z., Horsman, E., & Housen, B. (2011). Using vertical axis rotations to characterize off-fault deformation across the San Andreas fault system, central California. *Geology*, *39*(8), 711–714.

Withjack, M. O., Schlische, R. W., & Henza, A. A. (2007). Scaled experimental models of extension: dry sand vs. wet clay.

Reference From the Supporting Information

Bradley, D., & Roth, G. (2007). Adaptive thresholding using the integral image. Journal of Graphics Tools, 12(2), 13–21.

Cooke, M. L., & van der Elst, N. J. (2012). Rheologic testing of wet kaolin reveals frictional and bi-viscous behavior typical of crustal materials. Geophysical Research Letters, 39(1).

DeGroot, D., & Lunne, T. (2007). Measurement of Remoulded Shear Strength. Norwegian Geotechnical Institute. Report, 20061023–1.

Eisenstadt, G., & Sims, D. (2005). Evaluating sand and clay models: do rheological differences matter? Journal of Structural Geology, 27(8), 1399–1412.

Hatem, A. E., Cooke, M. L., & Toeneboehn, K. (2017). Strain localization and evolving kinematic efficiency of initiating strike-slip faults within wet kaolin experiments. Journal of Structural Geology. https://doi.org/10.1016/j.jsg.2017.06.011

Toeneboehn, K., Cooke, M. L., Bemis, S. P., Fendick, A. M., & Benowitz, J. (2018). Stereovision combined with particle tracking velocimetry reveals advection and uplift within a restraining bend simulating the Denali fault. Frontiers in Earth Science, 6, 152.