Prediction of microscale droplet instability in concentrated emulsion

DAVIS HOFFMAN
dwhoff@stanford.edu

December 15, 2017

I. INTRODUCTION

Understanding of microfluidics plays an important role in improving technologies in cell biology, biochemistry, electronics, and many other applications. One such application involves an emulsion of microscale droplets containing minute amounts of chemical reagents and/or biological samples. By passing the droplets through a microchannel, the contents of each can be probed sequentially. In practice, the rate at which droplets are processed should be as high as possible, as this determines the total processing time. However, as the throughput of droplets is increased, so is its susceptibility to break-up due to hydrodynamic instability. The phenomenon in the stable and unstable case is shown in Fig. 1.

Fig. 1. Stable and unstable case for two isolated droplets [2].

In previous studies, droplet break-up is seen to be dependent on the capillary number \(Ca\) and the critical offset \(\Delta x\), although the demarcation between the stable and unstable region is not well defined by these parameters [2]. In this bi-stable region, it seems that the process of droplet break-up is stochastic in nature, and the error of prediction is high. However, one promising way to improve the accuracy of prediction is to construct a more complex model and introduce additional features into the model, such as droplet interactions upstream of the constriction. In this study, convolutional neural networks (CNN) are used to uncover underlying droplet structures and interactions in order to classify future break-up.

II. DATA

The available data consists of 1,597 high-speed videos consisting of 30-40 frames each. Each video is provided with its class label as well as the critical offset \(\Delta x\). All videos have been recorded at the same flow rate. In order to make the problem more tractable, only videos of the same \(\Delta x\) are trained for a given model. Hence only a subset of the available data, roughly 660 videos, are used.

Preprocessing of the data is required in order to train a more meaningful model and for the algorithms to run more efficiently. Since the prediction should be based on what happens upstream of the channel and not within the channel, the frames are cropped to remove the channel portion of uniform cross section. In preparation for the deep learning stage, a square frame size is desired, and so padding is added as needed. A before-and-after example of image preprocessing is shown in Fig. 2.

Samples of processed images from both the stable and unstable case are grouped together...
Fig. 2. Example image preprocessed for use in classification algorithms.

Fig. 3. Sample of stable and unstable images before leading droplet breaks.

Fig. 4. Extracting relative position vectors between first 5 droplets.

III. Methods

Two approaches to image classification are taken for this problem: (1) non-deep learning techniques using feature extraction and (2) deep learning on the full image dataset. The deep learning approach is expected to yield a higher accuracy model. However, several non-deep learning approaches are explored which serve as a baseline for comparison. In addition, clustering and feature importance metrics are evaluated in order to understand the physical causes of the instability.

i. Feature Extraction

The goal of feature extraction in this problem is to identify underlying geometries in order to predict droplet break-up without using pixel-by-pixel data. Relevant properties may be calculated using pixel connectivity algorithms, which identifies each separate droplet region and fits the boundary with an ellipse. Every ellipse is then defined in terms of a spatial coordinate, major axis length, minor axis length, and azimuthal orientation. Fig. 4 details the process of extracting relative position vectors for the first five droplets in a frame.

The frames are typically truncated on the left side so that the trailing droplets are not fully captured. Consequently, the ellipse properties calculated from the image for these droplets cannot be used as is. However, the centroids may be corrected by shifting them an appropriate amount, assuming the truncated droplet is a perfect circle with a known area. The remaining properties of the truncated droplets cannot be reasonably estimated, and so are left out of the feature set. The total number of features considered for the non-deep learning models amounts to 14 (i.e. 8 relative position...
coordinates, 2 major axis lengths, 2 minor axis lengths, and 2 azimuthal orientations).

ii. Deep Learning

A CNN model for the MNIST dataset in Keras has been adapted for the purpose of image classification of droplet break-up. The model architecture is shown in Table 1. The MNIST dataset consists of 70,000 images of size 28x28. Since the droplet frame size is much larger than the MNIST frame size, the convolutional layer is trained on the same MNIST dataset, but scaled up to 42x42. After the CNN learns the rescaled MNIST dataset with over 99% accuracy, the convolutional layer weights are fixed, and the fully connected layers are retrained on the droplet datasets. An example frame from the droplet dataset rescaled to 42x42 from original size is shown in Fig. 5.

Table 1: CNN architecture for MNIST classifier.

<table>
<thead>
<tr>
<th>Layer (type)</th>
<th>Output Shape</th>
<th>Params #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conv1 (Conv2D)</td>
<td>(None, 30, 30, 30)</td>
<td>7800</td>
</tr>
<tr>
<td>MP1 (MaxPooling2D)</td>
<td>(None, 30, 15, 15)</td>
<td>0</td>
</tr>
<tr>
<td>Conv2 (Conv2D)</td>
<td>(None, 15, 17, 17)</td>
<td>49065</td>
</tr>
<tr>
<td>MP2 (MaxPooling2D)</td>
<td>(None, 15, 8, 8)</td>
<td>0</td>
</tr>
<tr>
<td>Dropout</td>
<td>(None, 15, 8, 8)</td>
<td>0</td>
</tr>
<tr>
<td>Flatten</td>
<td>(None, 950)</td>
<td>0</td>
</tr>
<tr>
<td>FC1 (Dense)</td>
<td>(None, 128)</td>
<td>123000</td>
</tr>
<tr>
<td>FC2 (Dense)</td>
<td>(None, 50)</td>
<td>6450</td>
</tr>
<tr>
<td>S0 (Dense)</td>
<td>(None, 2)</td>
<td>182</td>
</tr>
<tr>
<td>Total params</td>
<td>134,445</td>
<td></td>
</tr>
<tr>
<td>Trainable params</td>
<td>129,508</td>
<td></td>
</tr>
<tr>
<td>Non-trainable pars</td>
<td>4,845</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 5. Example dataset scaled from 151x151 to 42x42 for deep learning model.

Fig. 6. Learning curve for (a) GDA, (b) logistic regression, and (c) SVM using Gaussian kernel.

IV. Results

Non-deep learning models using extracted features demonstrate a significant ability to predict droplet break-up. Datasets corresponding to a nominal \( \Delta x = 13 \) pixels were used \((11 \leq \Delta x \leq 15 \) was actually used in order to increase the number of training examples). The proportion of training and testing sets used is 70:30%.

The learning curve for Gaussian Discriminant Analysis (GDA), logistic regression, and support vector machine (SVM) averaged over 20 ensembles, is shown in Fig. 6. The results for GDA suggest that little improvement beyond 30% test error is achievable. This is reasonable, since the features are highly dependent on each other. For this reason, logistic regression is expected to be a better model for the data. Logistic regression indeed performs slightly better, with lower variance. SVM also performs even better, but the learning curve suggest increasing the number of training examples to reduce variance even further.

Random forest proved to be a good classifier given the extracted feature set. The aggregation of multiple decision trees results in a model that improves generalization by reducing overfitting. The number of trees is chosen to be 50, based on the error plots shown in Fig. 7, although any choice above 15 would be
reasonably accurate.

![Fig. 7. RF test error versus number of ensembled trees.](image1)

A summary of all the model accuracies performed using the feature extraction dataset is given in Table 2.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDA</td>
<td>68.7</td>
</tr>
<tr>
<td>Log. Reg.</td>
<td>70.1</td>
</tr>
<tr>
<td>SVM</td>
<td>87.1</td>
</tr>
<tr>
<td>RF</td>
<td>85.1</td>
</tr>
</tbody>
</table>

Table 2: Final test accuracy of non-deep learning models

Results from the previous models serve well as a benchmark for more complex neural network models. Assuming the training set size is sufficiently large, these models should outperform the models on the feature extraction dataset. A major concept behind neural networks is to automatically learn the important features within the hidden layers. Therefore, the previous results are a lower bound to the performance of a well constructed neural network model.

Initially, a simple neural network with one fully connected layer is trained on the vectorized pixel values directly. The hidden layer is composed of 300 neurons, cross entropy defined as the loss, and L2 regularization on the scaling weights. The training history is shown in Fig. 8.

The CNN model proved effective in reaching classification accuracies above the baseline values. The training history for the pretrained MNIST model is shown in Fig. 9. The model trains relatively quickly, as only the weights of the fully connected layers are updated during back propagation.

![Fig. 8. Accuracy over the training history for simple fully connected neural network.](image2)

V. DISCUSSION

This machine learning study was successful in developing microdroplet instability classifiers using images prior to the breakup event. However, much work can be done in order to improve the robustness of these models. Surprisingly, the performance of the CNN with so few images in the training set is already quite impressive. Even for an MNIST classifier it is difficult to achieve accuracies above 90% with
several thousand training images.

It remains a challenge to understand and quantify the physical mechanism that occurs in the bistable flow regime. One way to see the effect of individual features on the model’s discriminative ability is to calculate variable feature importance using decision trees, as shown in Fig. 10.

![Feature importance](image)

**Fig. 10.** Variable feature importance from random forest classifier.

According to this analysis, the most important features are those corresponding to the leading droplet pair. The remaining features corresponding to the trailing droplets are of secondary importance, but by no means should be left out. These features may require more accurate measurement when being extracted from the frames. Larger frames which capture the full set of trailing droplets would help in extracting these features. This would also allow additional features from trailing droplets to be added to the feature set, such as their elongation length and orientation.

Unsupervised learning algorithms, like k-means clustering, are useful for qualitative analysis of the unstable examples. Clusters are identified according to the features extracted from the image, and the corresponding pictures for each may be separated for visual comparison. Two example clusters are shown in Fig. 11. The appropriate number of clusters to use is not clear, although redundant clusters seem to appear around $k = 10$. For a given cluster, there is often a second cluster which is its symmetric counterpart, which allows for the estimation $5 \leq k \leq 10$.

![Clusters](image)

**Fig. 11.** Example clusters for the unstable case, identifying unfavorable droplet stretching and packing pattern.

An important aspect not considered in this study is the time evolution of the droplets leading up to the instability. By considering the time encoded information, relative velocities of the droplets may be added to the feature set. These features may have high importance, as relative velocities indicate in what manner the droplets will enter the channel at a later time.

One possible way to incorporate time series data into the existing CNN model is to stack two additional previous frames with the existing frame. This allows for the use of the full three dimensional data structure that CNNs are able to process, since each frame by itself is only two dimensional.

### VI. Conclusions

Baseline classifiers have been successfully applied to the binary classification problem of microdroplet break-up using feature extraction with accuracy up to 87.1%. Deep learning algorithms are currently trained to 90.0% accuracy, but more data may be required to optimize the weights, especially if transfer learning is the be bypassed.

### References

in python with keras. 2016.