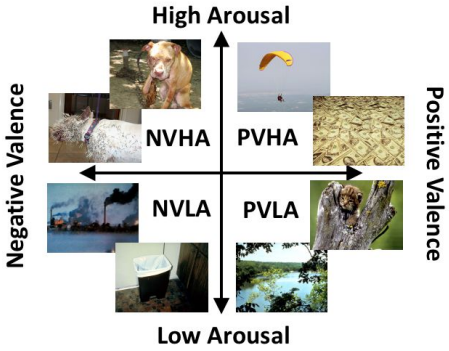
Interpretable Machine Learning Approach to Human Emotion Recognition and Visualization

Presenters: Vahan & Bin

Content

Introduction

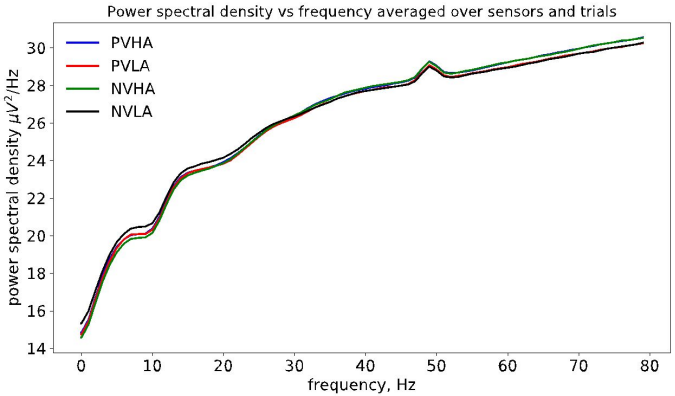
1. The electroencephalogram (EEG) has been a popular approach for examining brain activity.
2. For our experiment:
   - We use 59 electrodes to gain EEG emotional data.
   - We classify four emotions
     - NVHA
     - PVHA
     - NVLA
     - PVLA
   - We have applied SVM (85%) and CNN (81%) model to the EGG data.
3. But we do not understand how SVM and CNN model work on EEG data. EEG data is totally different from data like images. For images data, CNN can capture their textures , colors, lines, etc.

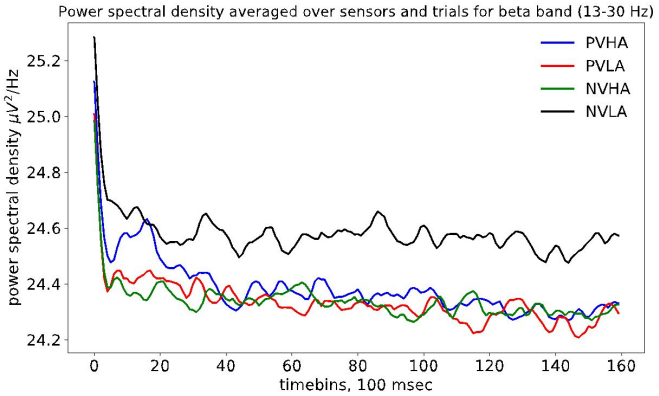


Introduction

4. EEG data visualizations

- NVHA
- PVHA
- NVLA
- PVLA



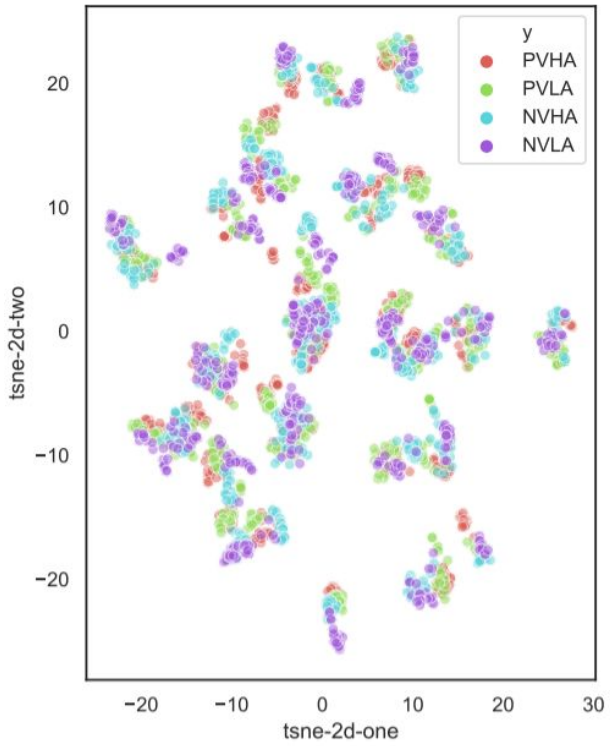


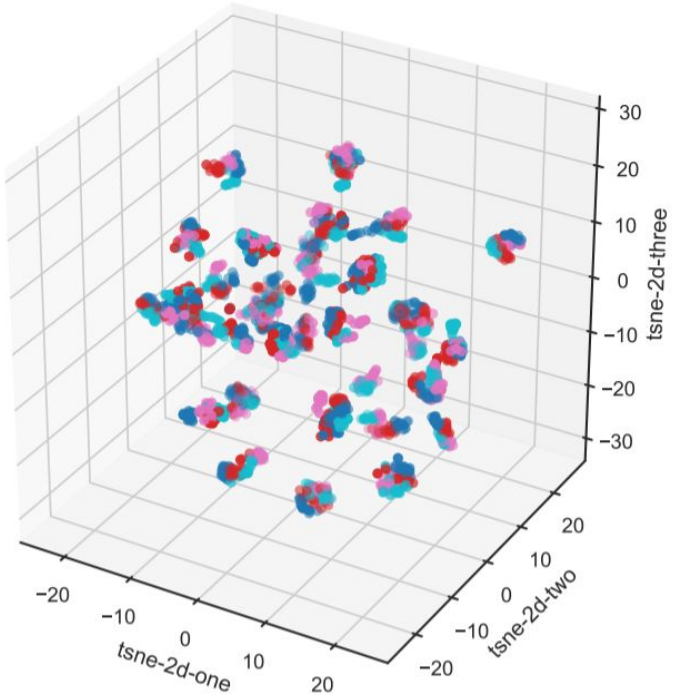
5. What we have done to understand the model and the data

- Visualize CNN layers and dense layers
- Visualize the SVM boundary
- Gain insight from the visualization and change CNN architecture

Data Visualization -- Input map

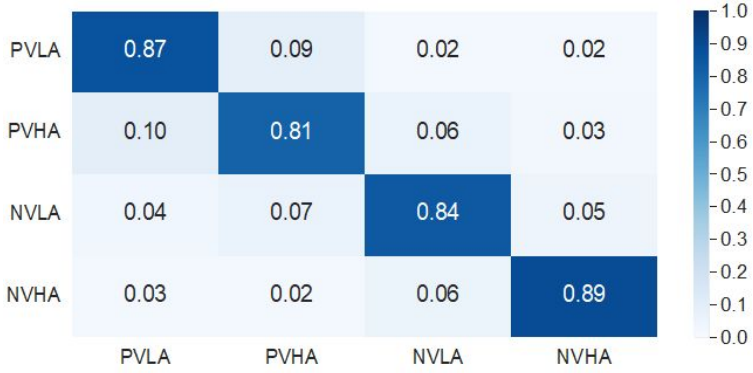
Input map



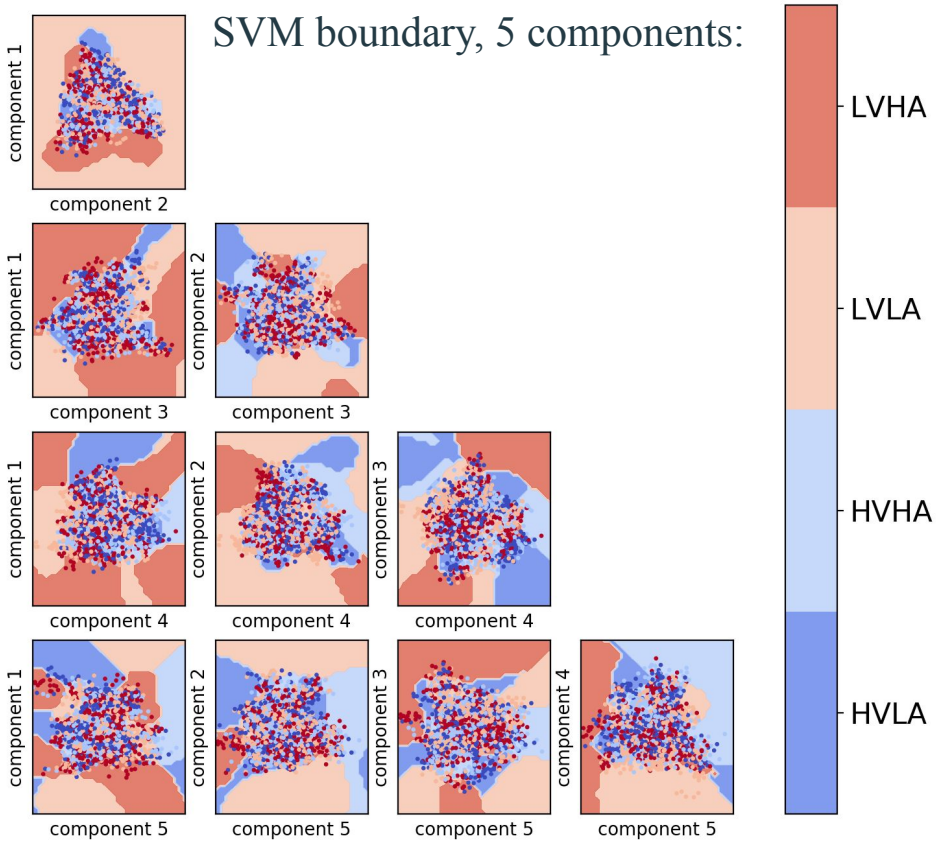


Data Visualization -- SVM

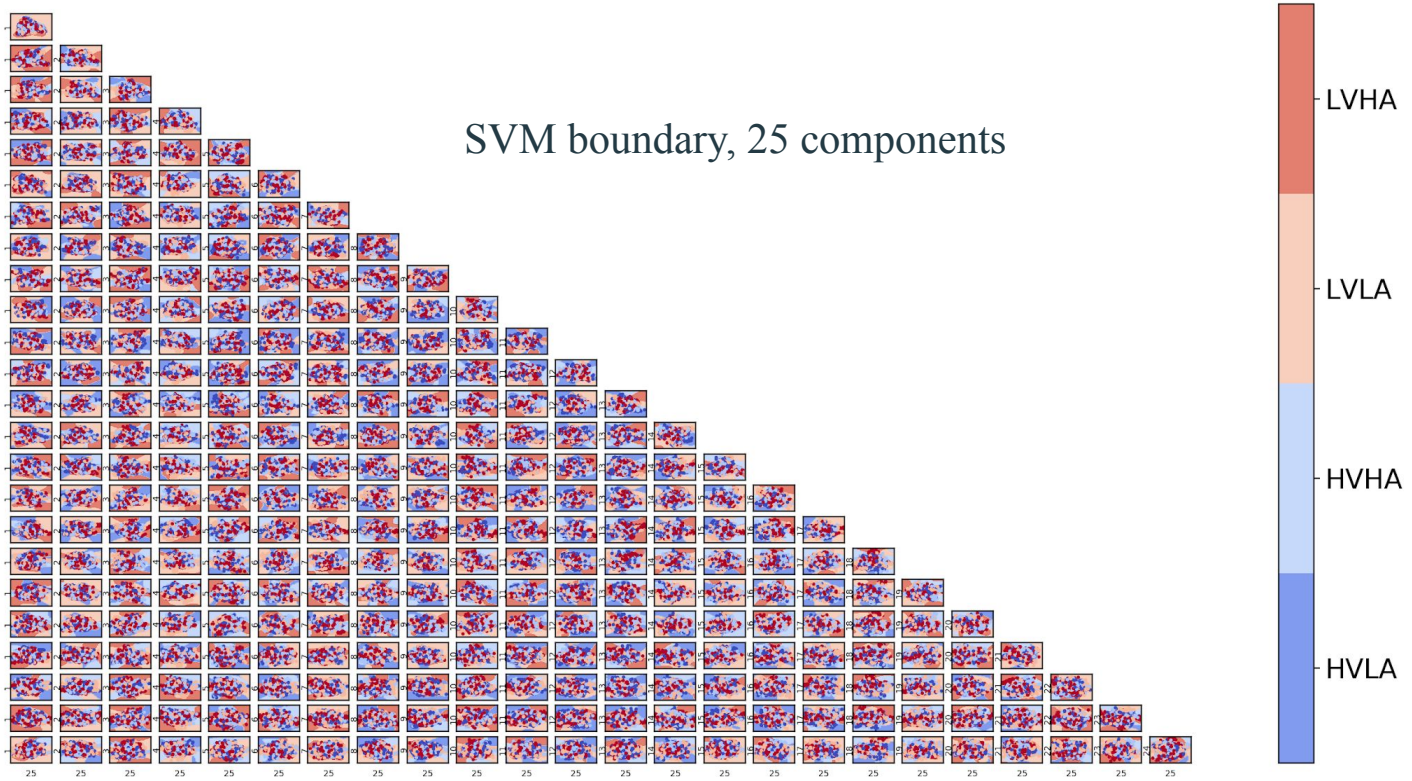
Confusion Matrix



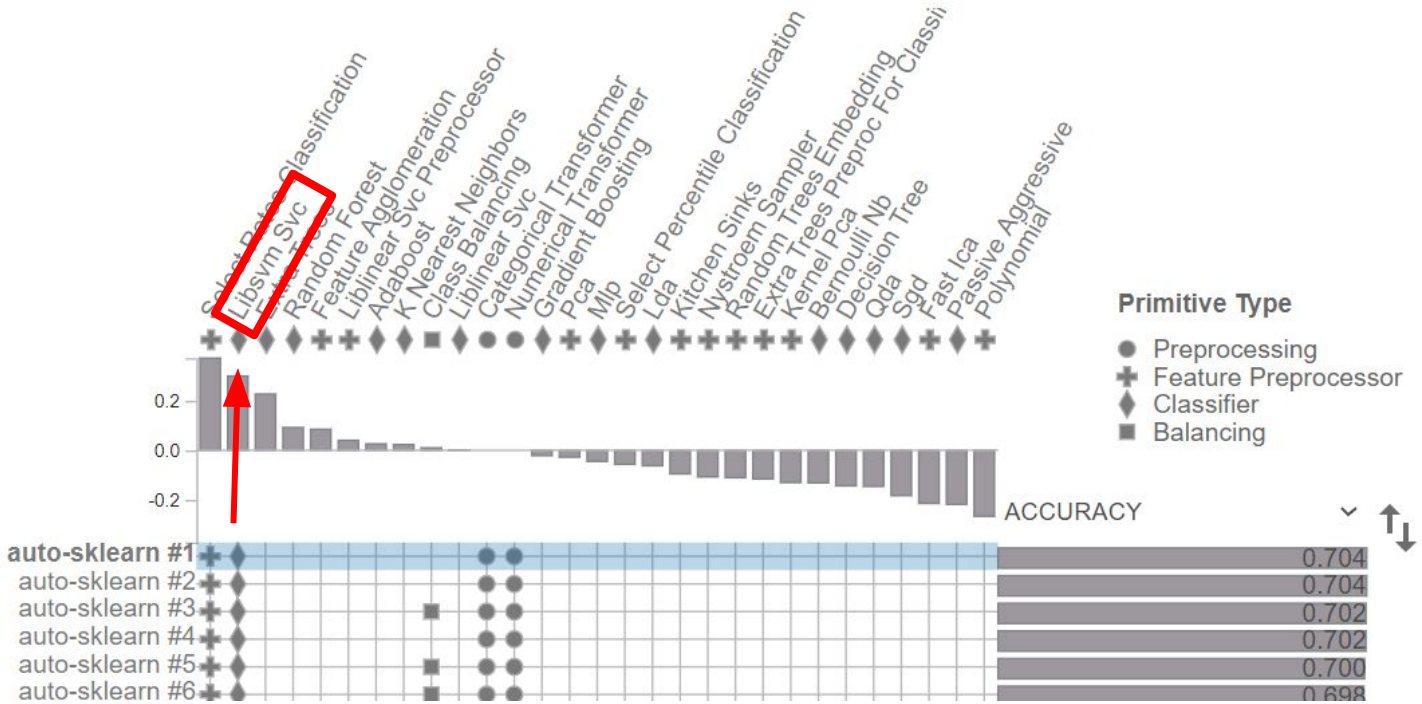
SVM boundary, 5 components:



Data Visualization -- SVM

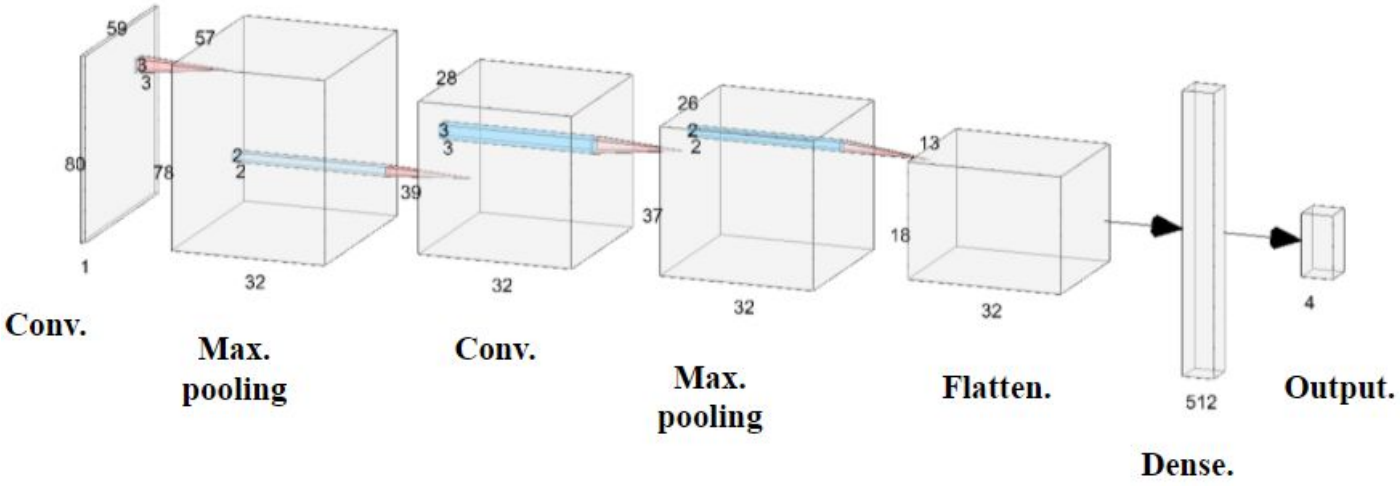


Data Visualization -- SVM-- AutoML



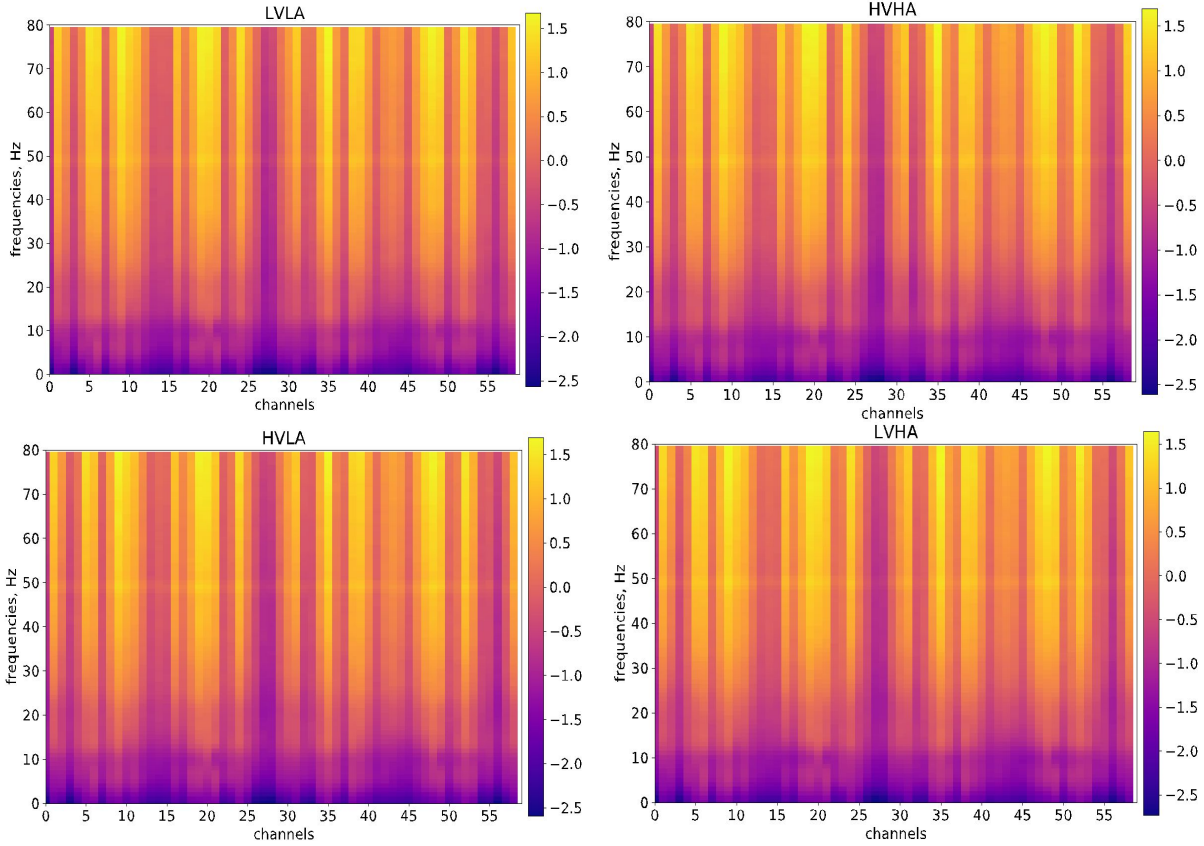
Data Visualization -- CNN Layers -- Old model

- Old Model architecture (81%)

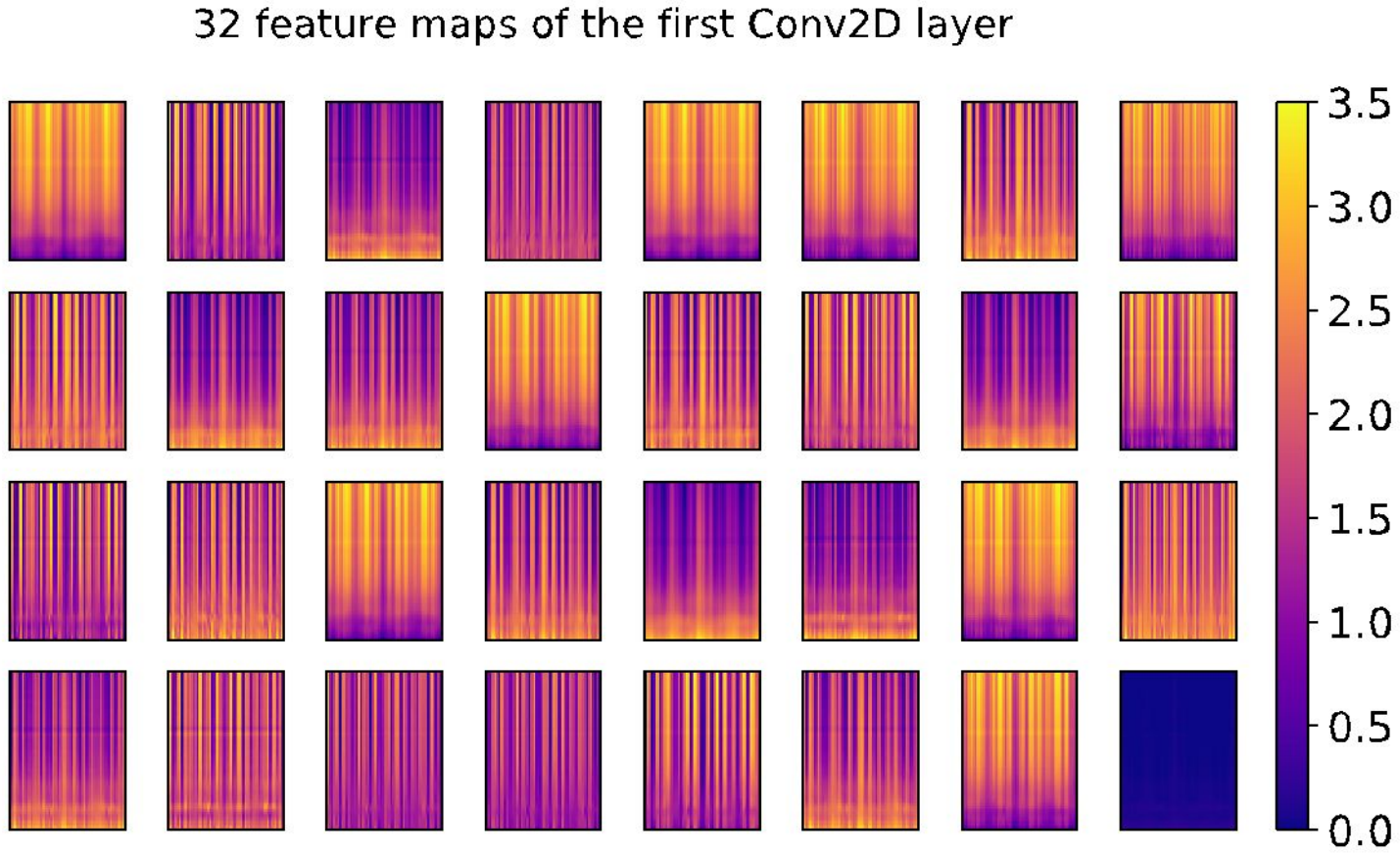


- Insight, why do we change to new CNN architecture?

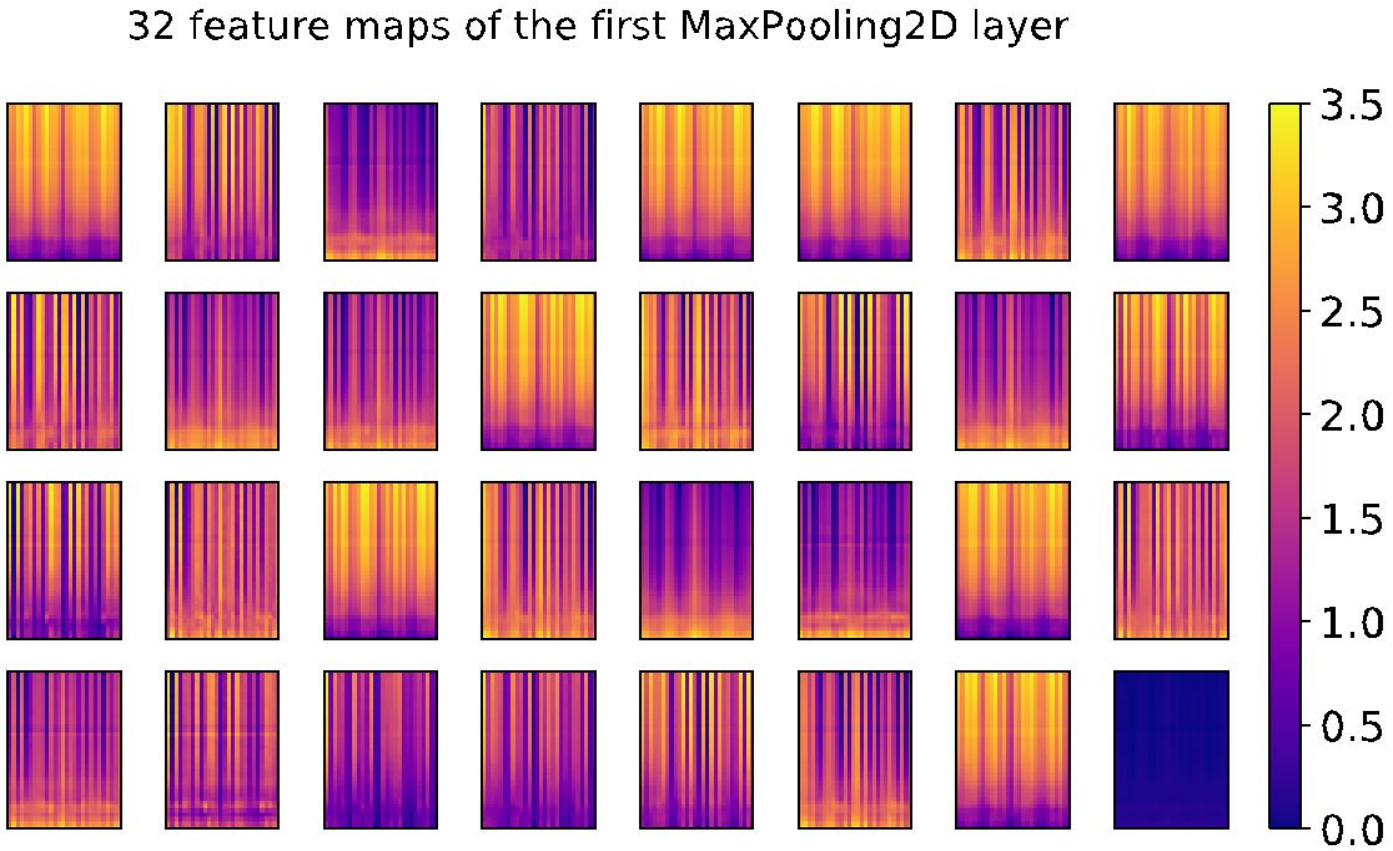
Spectrograms averaged over trials



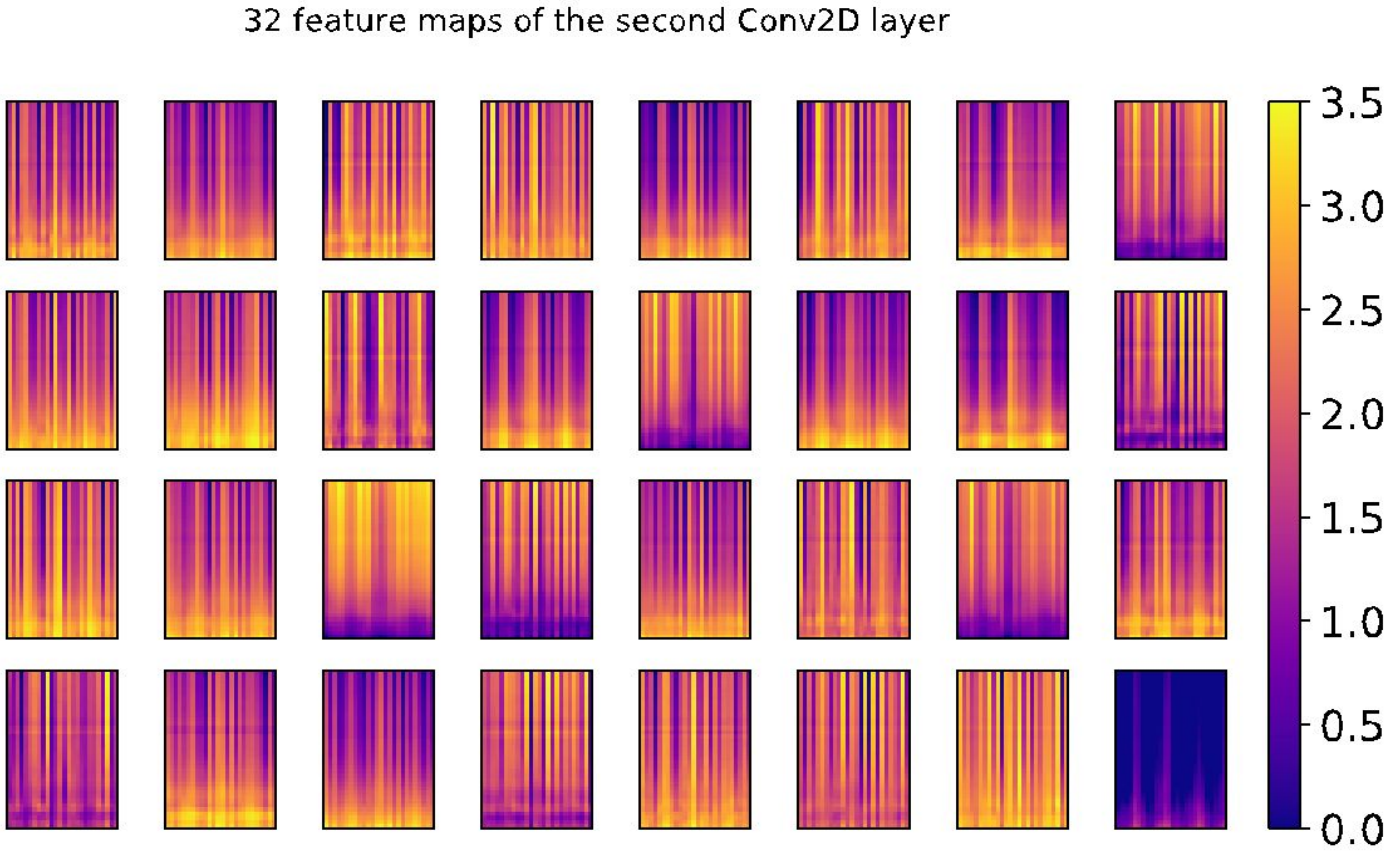
Feature Maps, Conv 1st layer



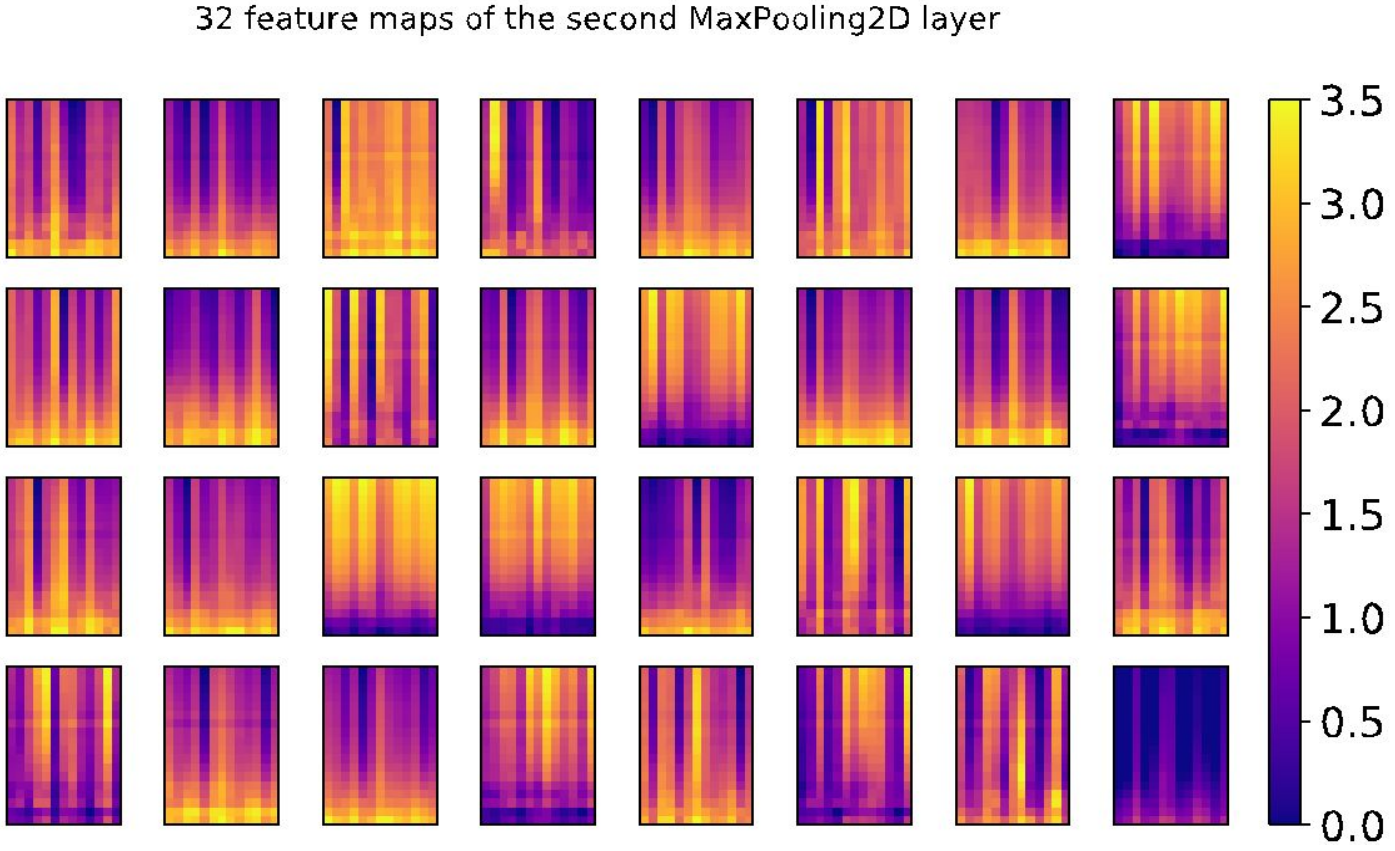
Feature Maps, MaxPooling 1st layer



Feature Maps, Conv 2nd layer

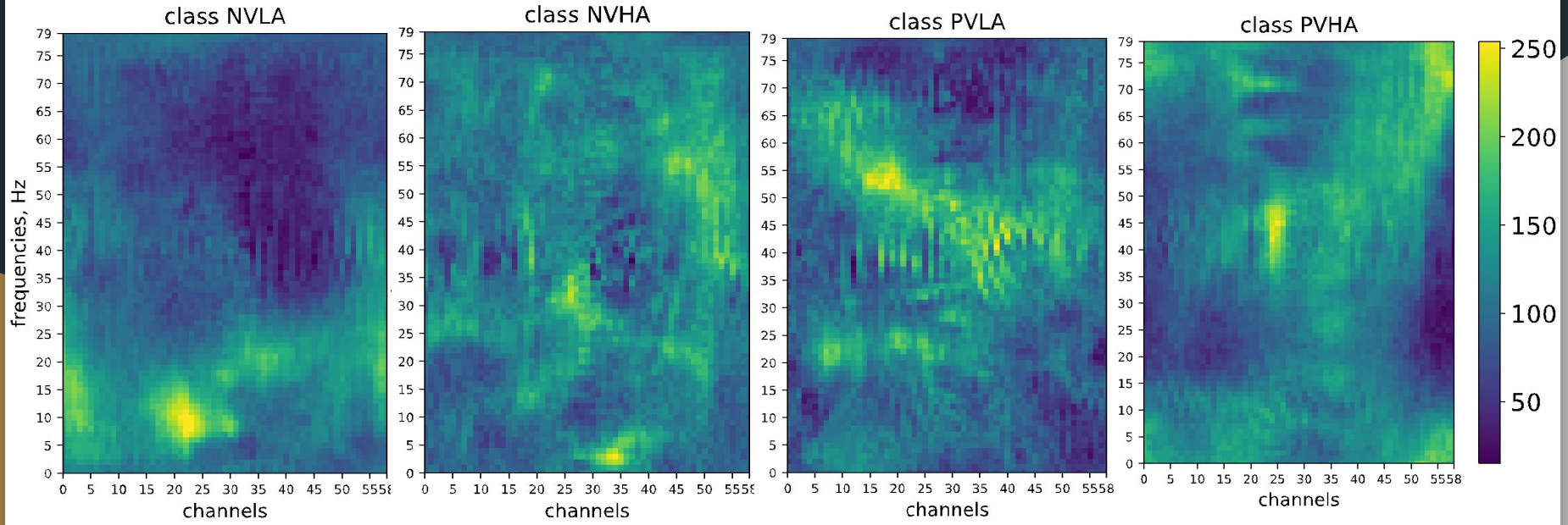


Feature Maps, MaxPooling 2nd layer

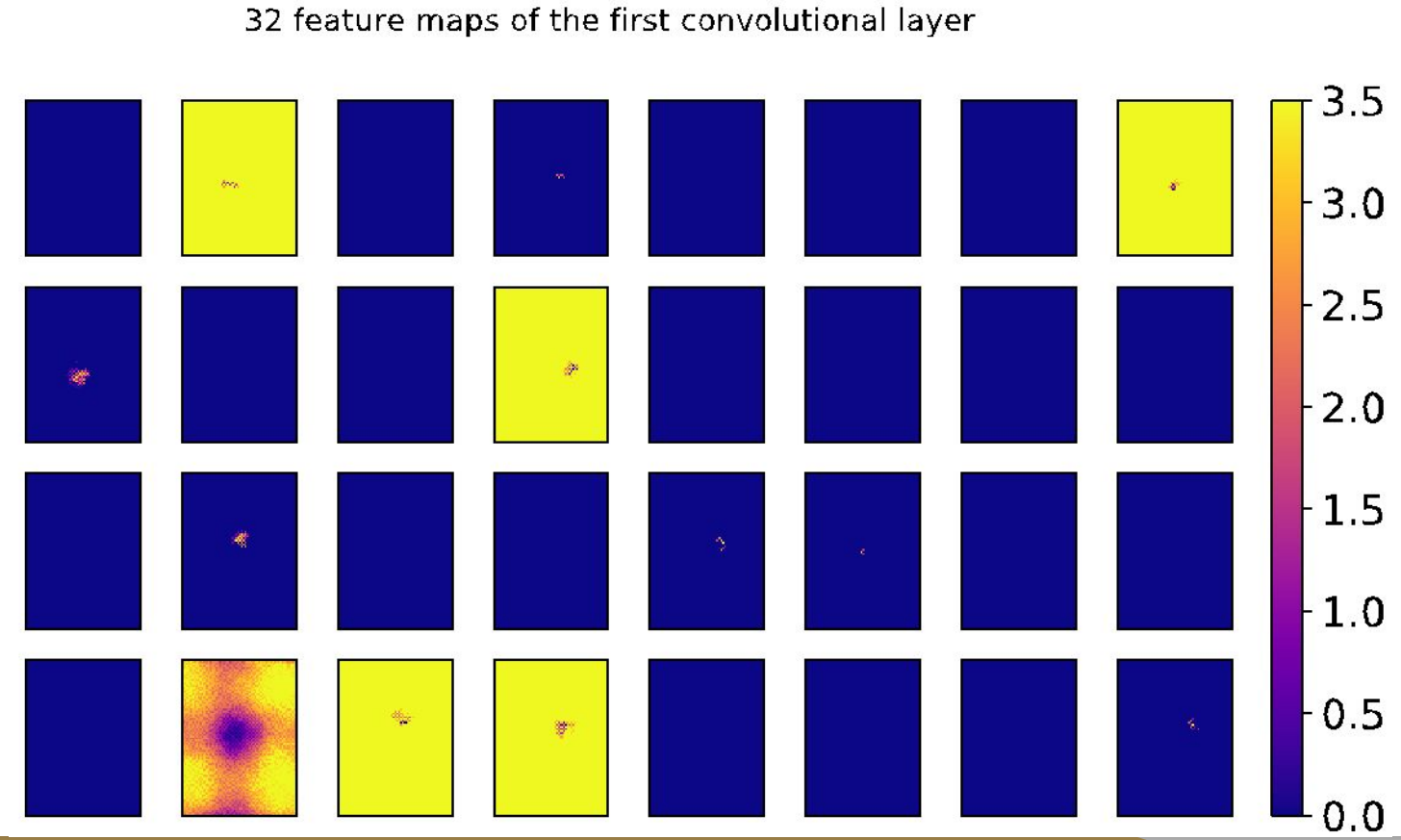


Visualization of CNN Dense Layer Activation Maximization

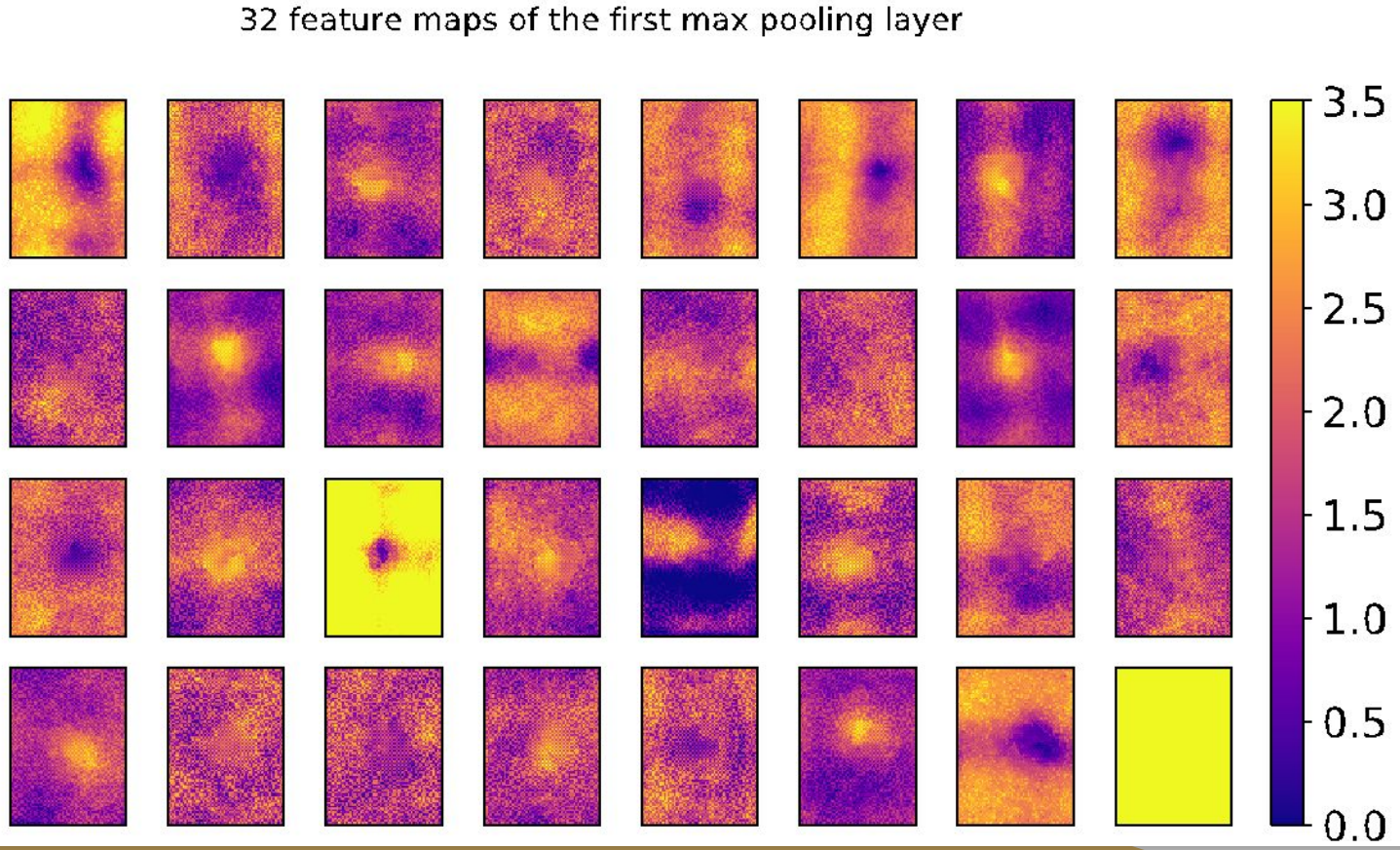
- Old Model



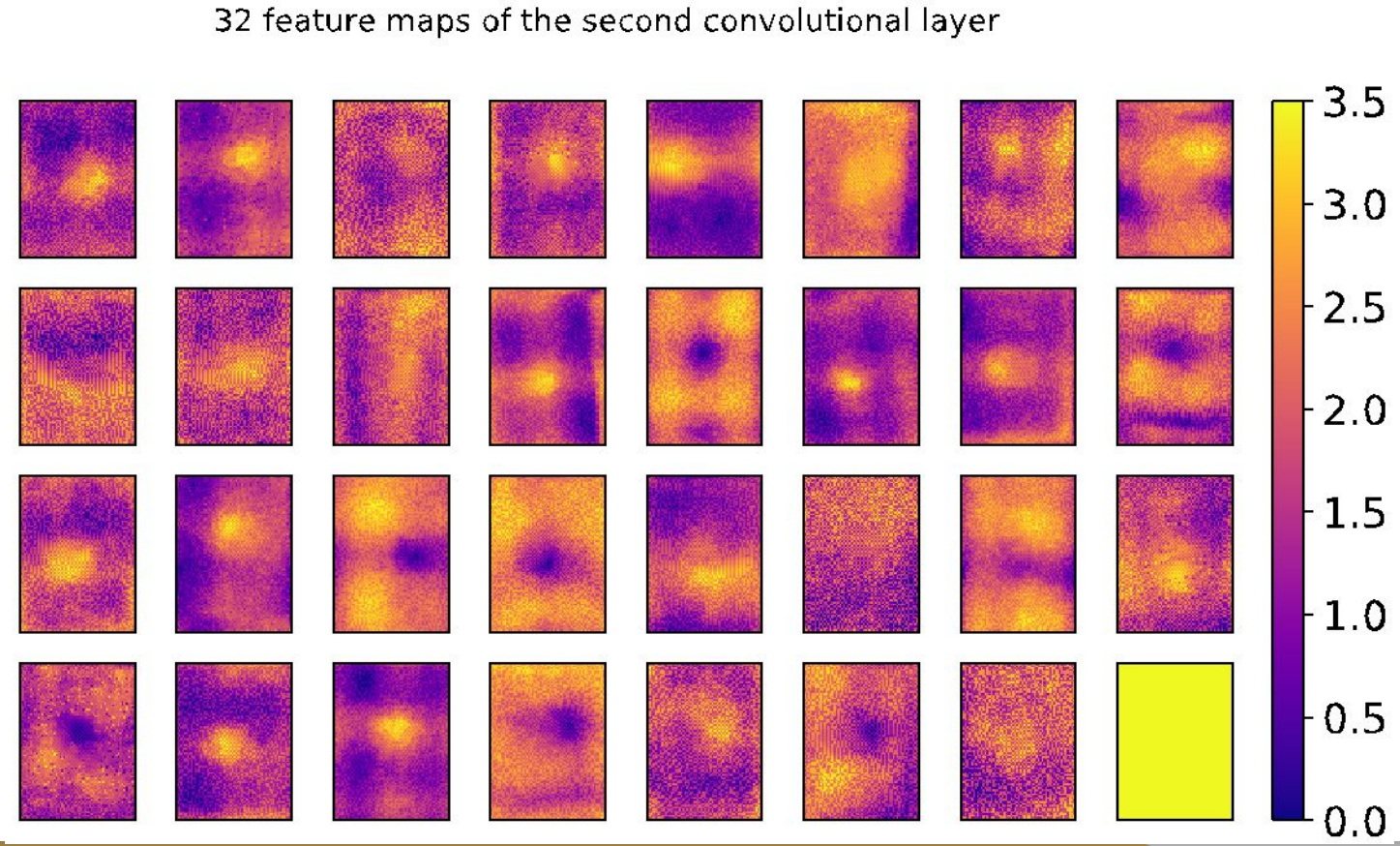
Visualizing Conv 1 layer Activation Maximization



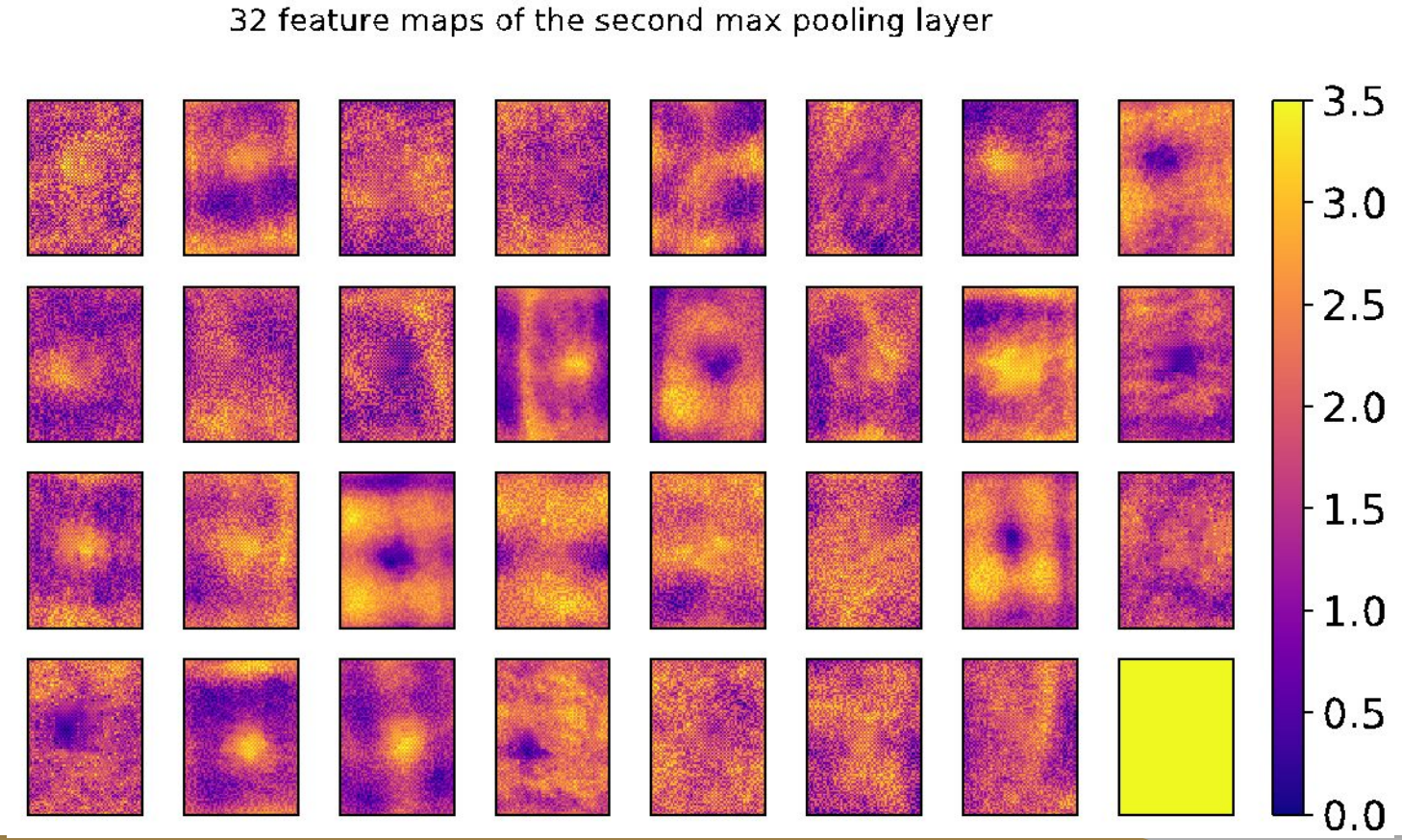
Visualizing MaxPooling 1 layer Activation Maximization



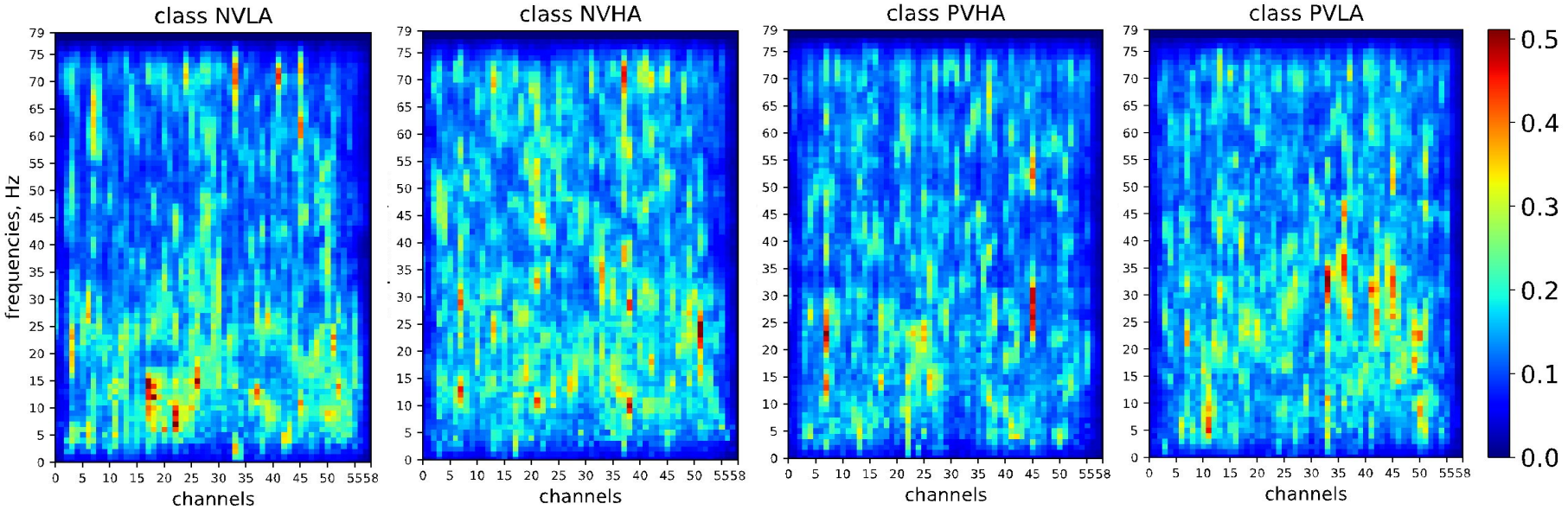
Visualizing Conv 2 layer Activation Maximization



Visualizing MaxPooling 2 layer Activation Maximization

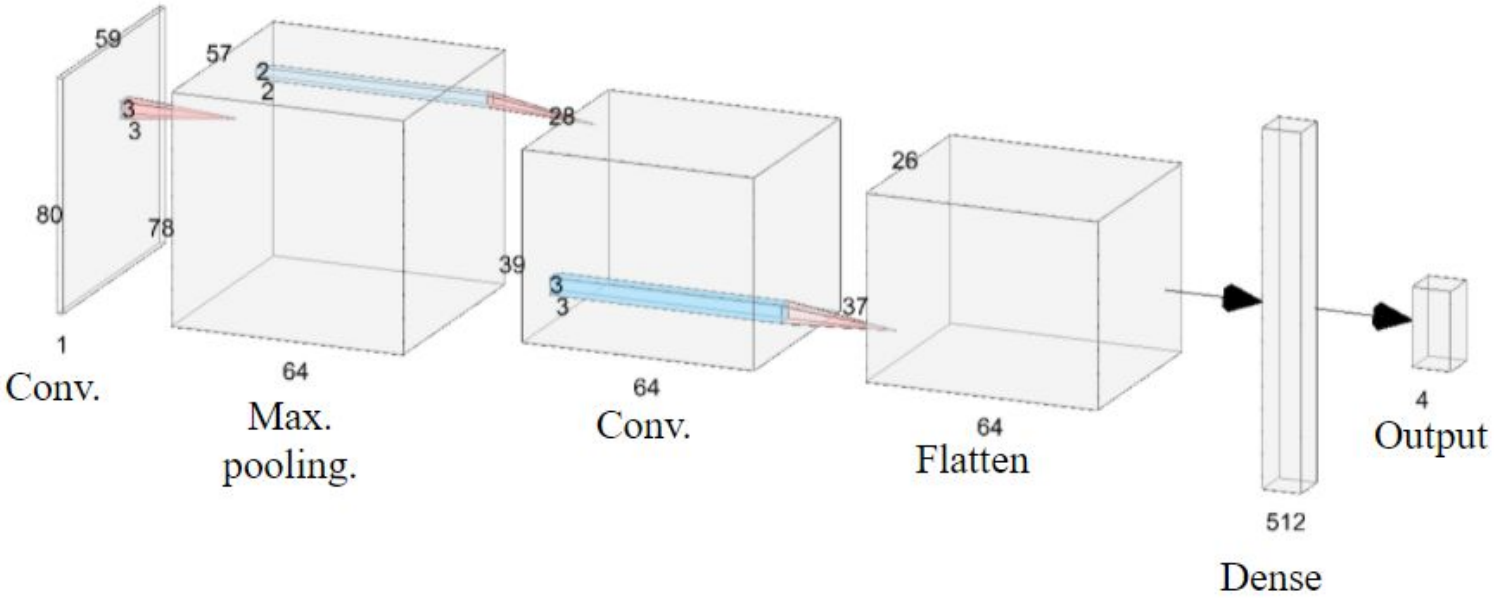


Saliency Maps, Old Model

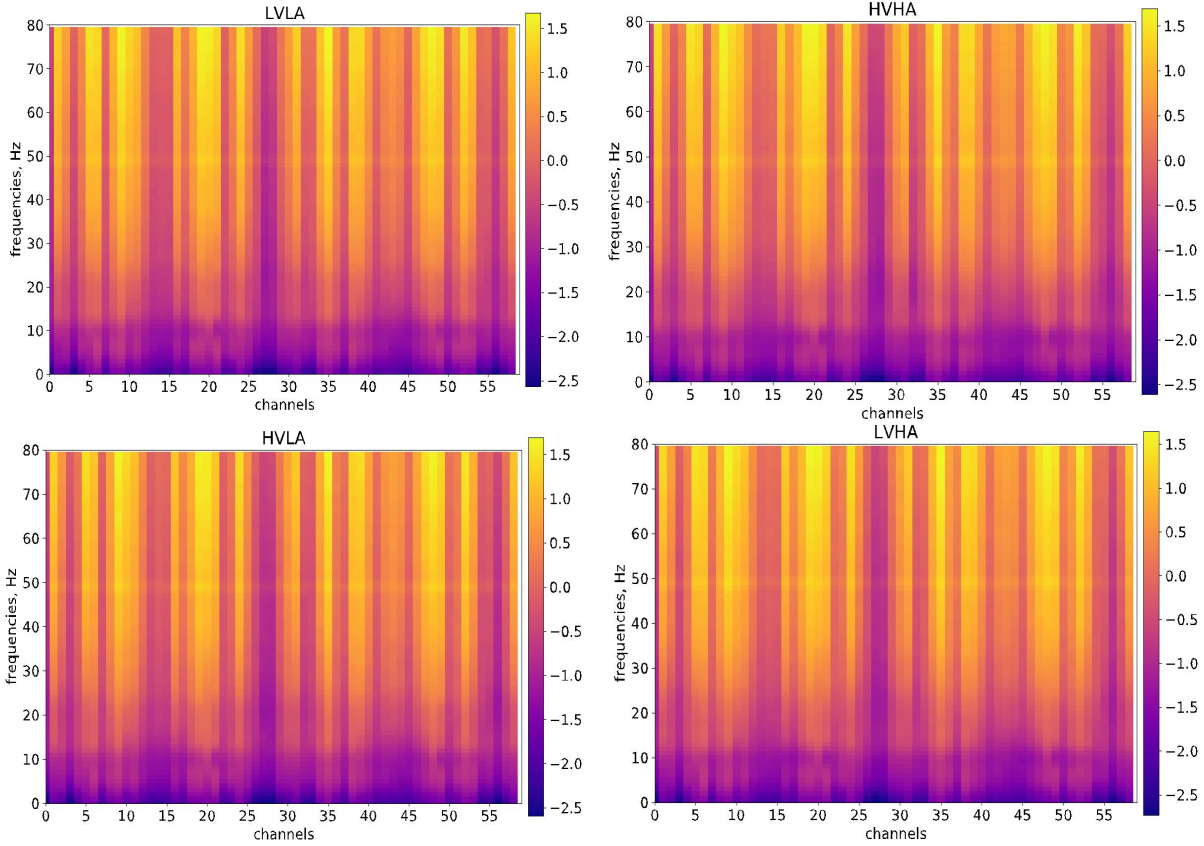


Data Visualization -- CNN Layers -- New model

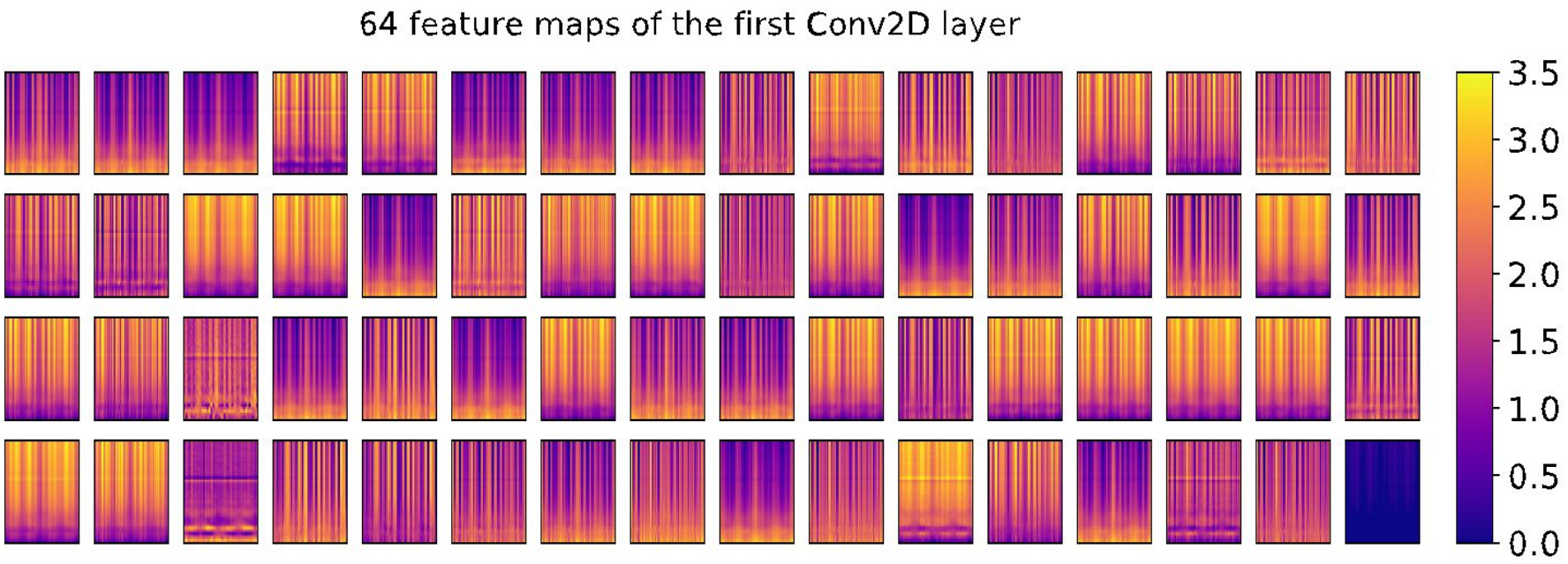
- New Model architecture (85%)



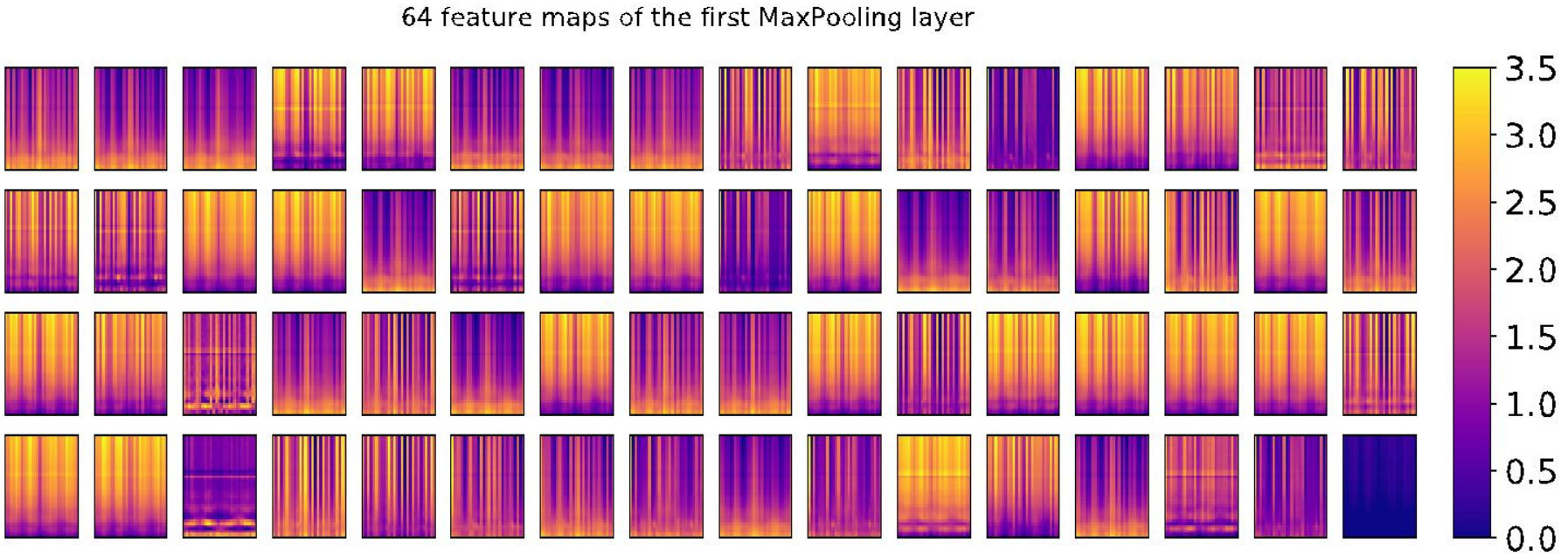
Spectrograms averaged over trials



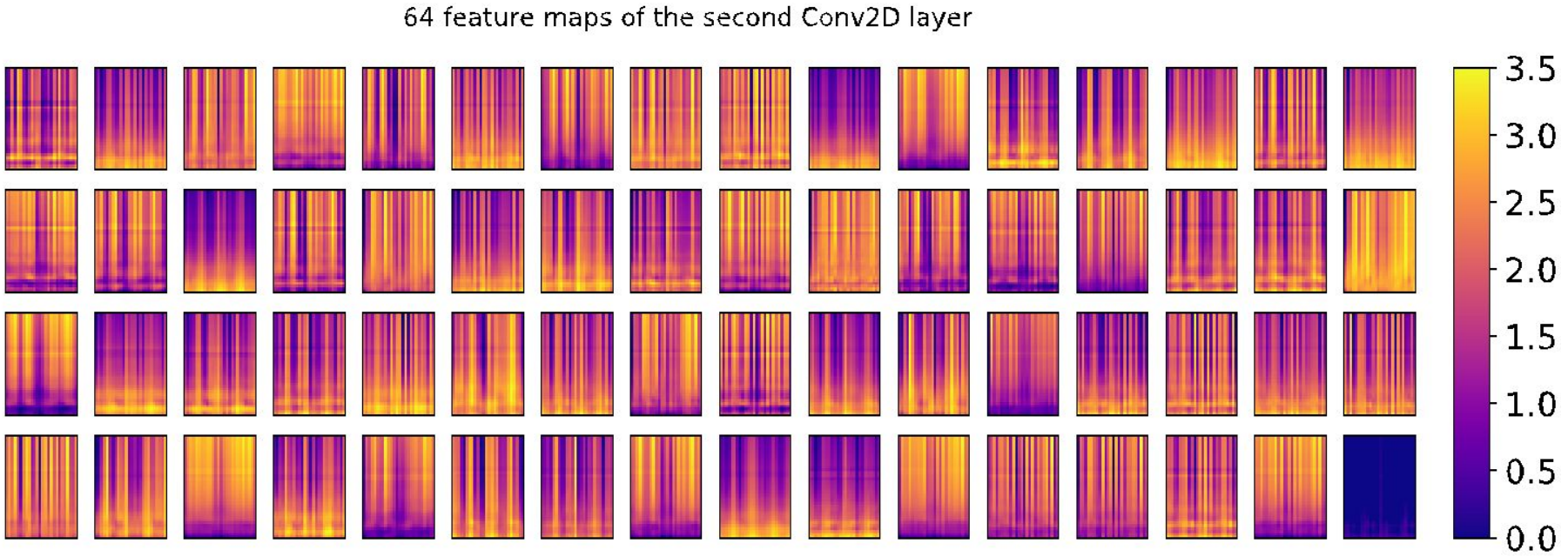
Feature Maps, Conv 1st layer



Feature Maps, MaxPooling 1st layer

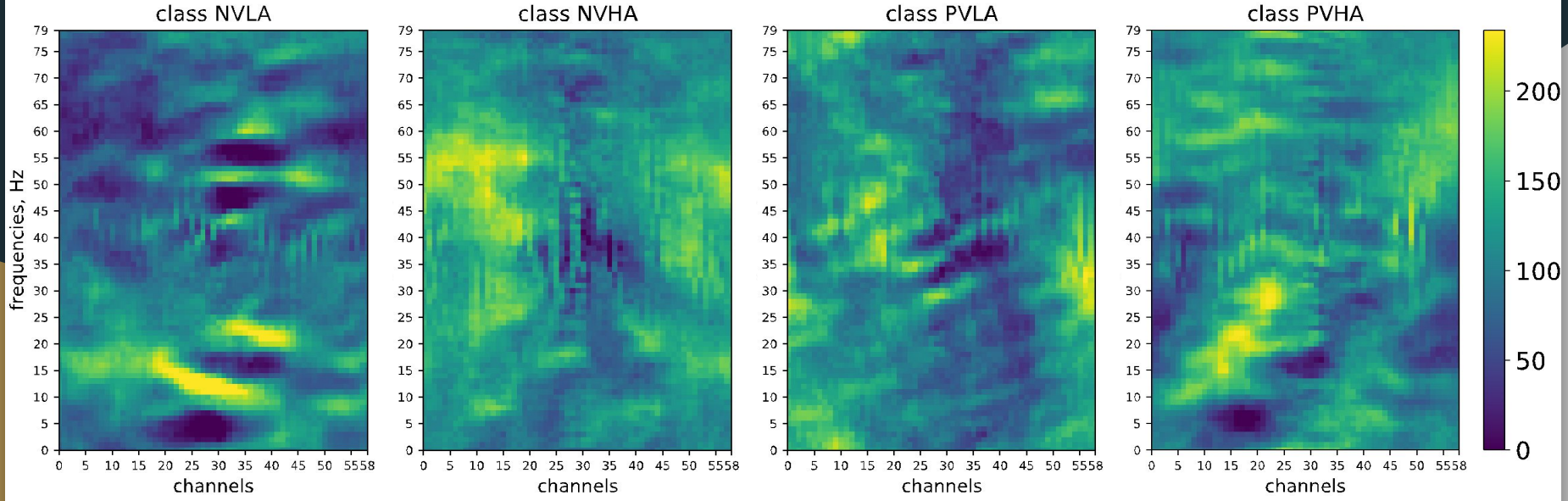


Feature Maps, Conv 2nd layer

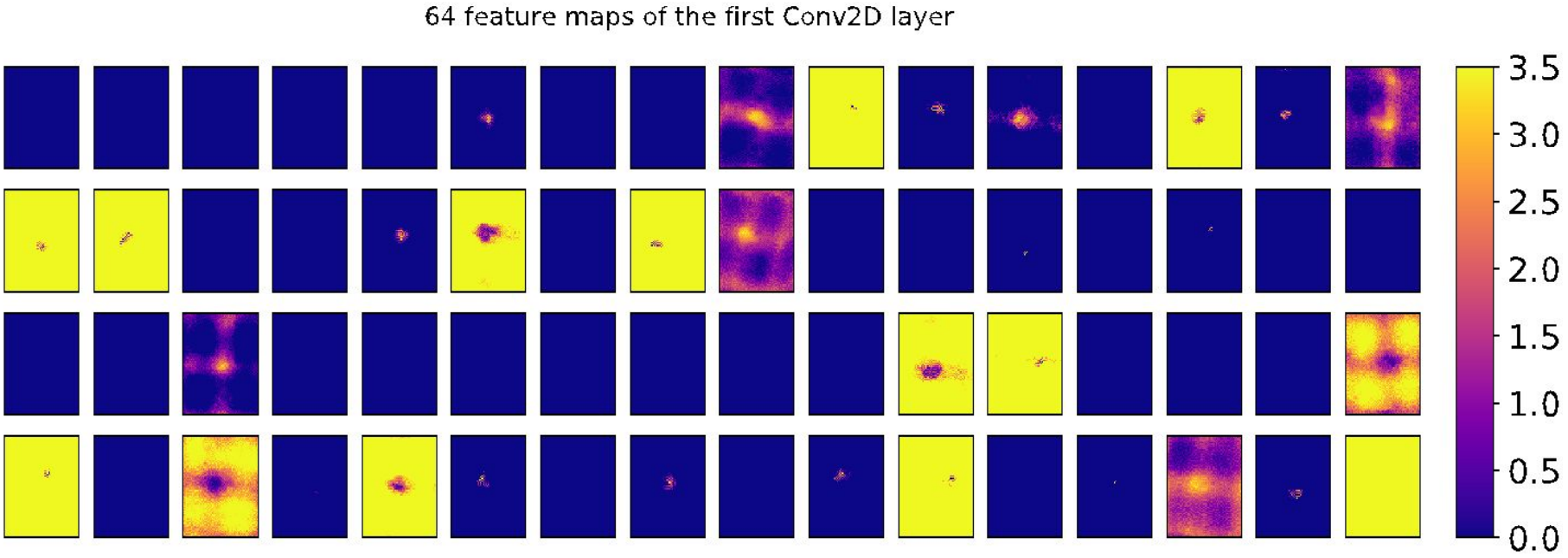


Visualization of CNN Dense Layer Activation Maximization

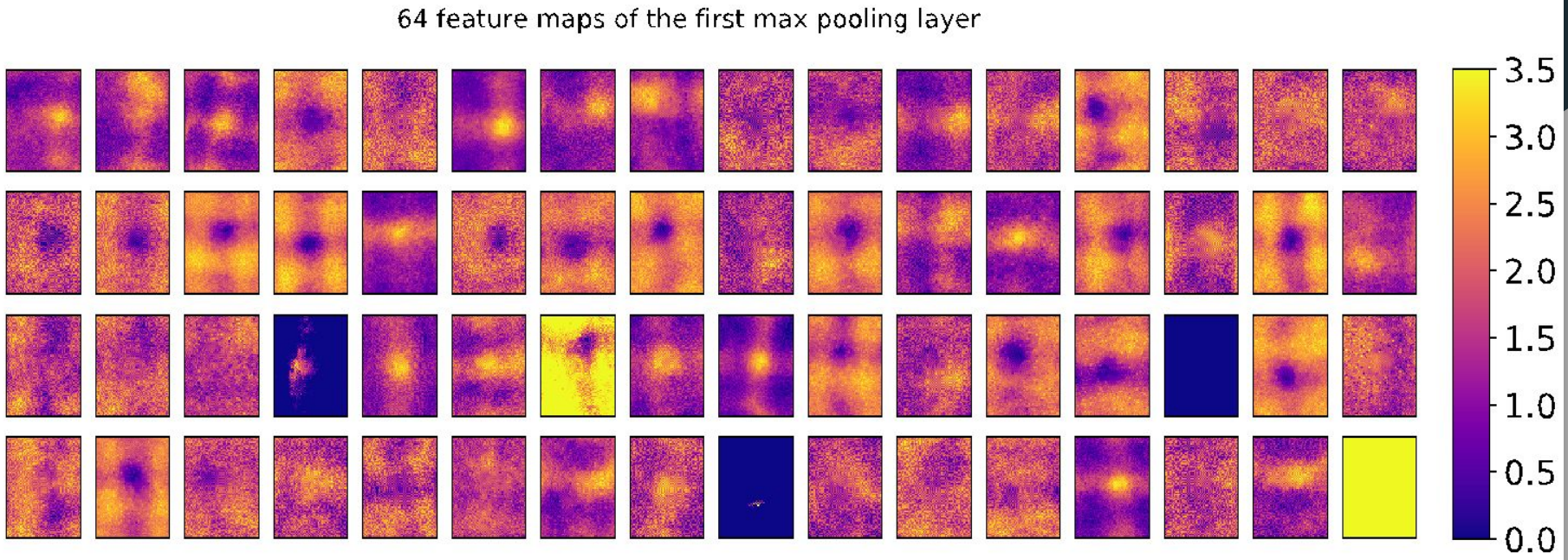
- New Model



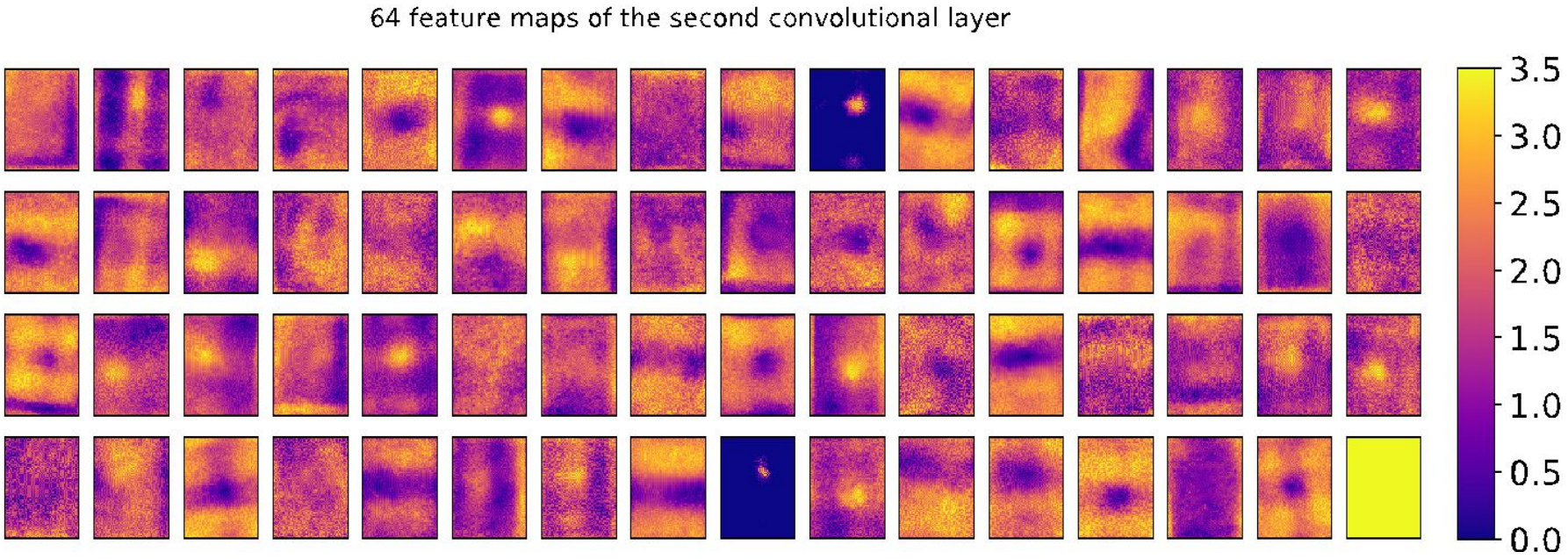
Visualizing Conv 1 layer Activation Maximization



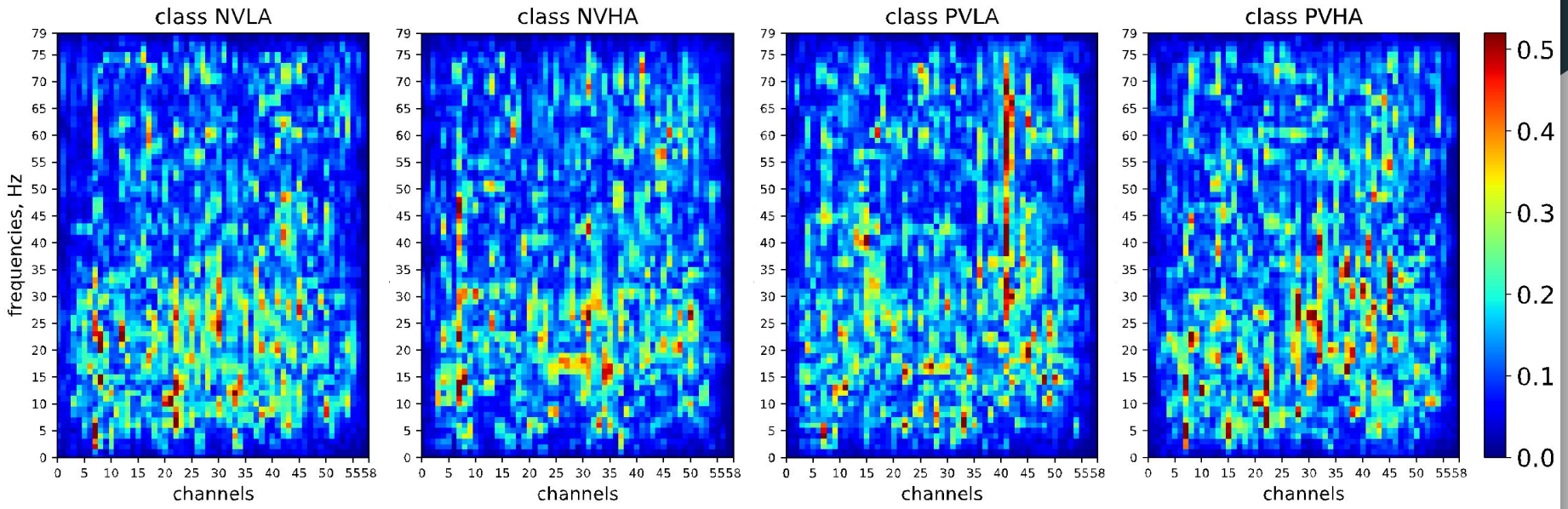
Visualizing MaxPooling 1 layer Activation Maximization



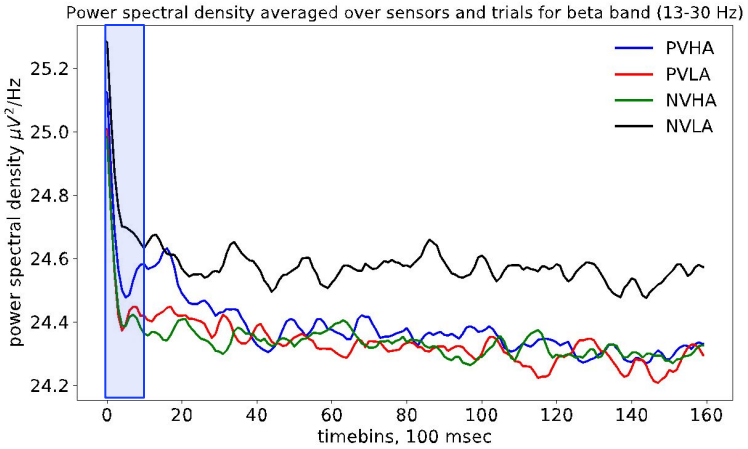
Visualizing MaxPooling 1 layer Activation Maximization



Saliency Maps, New Model



Proper data processing for old model



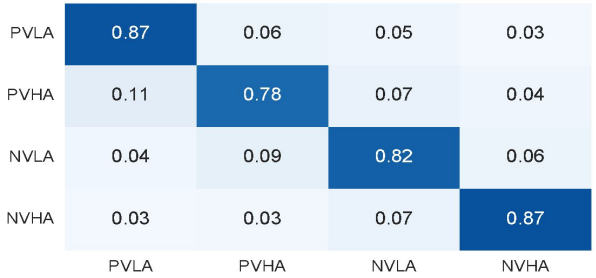
Old model without time trimmed:

- Average Accuracy: 0.8341327072502654
- Average Precision: 0.8303159213115249
- Average Recall: 0.8333851884010003
- Average F1: 0.8318477236985471

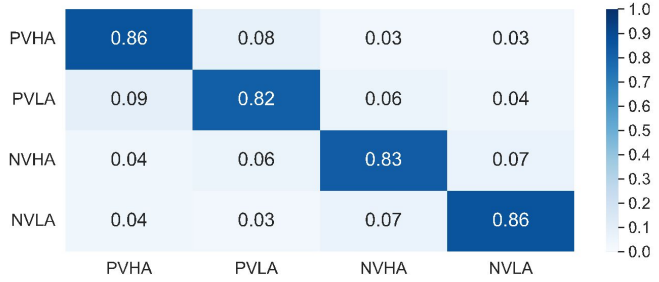
Old model with time trimmed from 10-th timebin (≈500 msec):

- Average Accuracy: 0.8426966292134831
- Average Precision: 0.8487601054053673
- Average Recall: 0.8466089497052871
- Average F1: 0.8476831628168261

Old model



Old model time trimmed



New model vs old model 10-fold CV

Old model without time trimmed:

- Average Accuracy: 0.8341327072502654
- Average Precision: 0.8303159213115249
- Average Recall: 0.8333851884010003
- Average F1: 0.8318477236985471

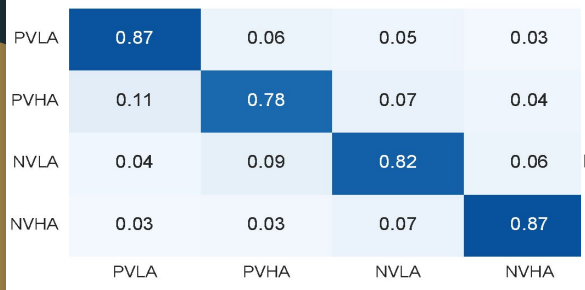
New model without time trimmed:

- Average Accuracy: 0.8431256638157526
- Average Precision: 0.8422079611526454
- Average Recall: 0.8409212020118906
- Average F1: 0.8409212020118906

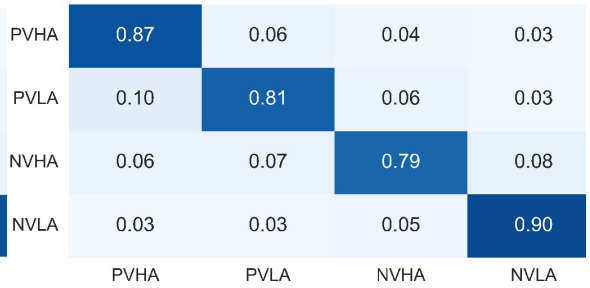
New model with time trimmed from 10-th timebin (≃500 msec):

- Average Accuracy: 0.8480015652076694
- Average Precision: 0.8487601054053673
- Average Recall: 0.8472898214772309
- Average F1: 0.8466089497052871

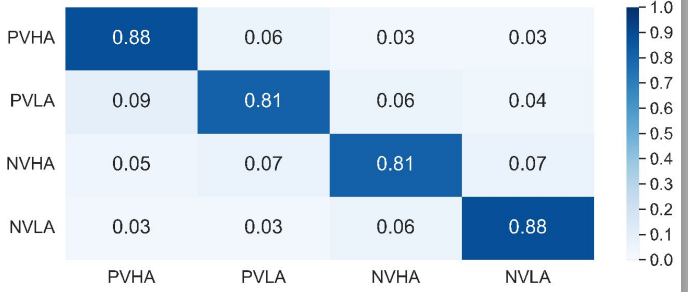
Old model



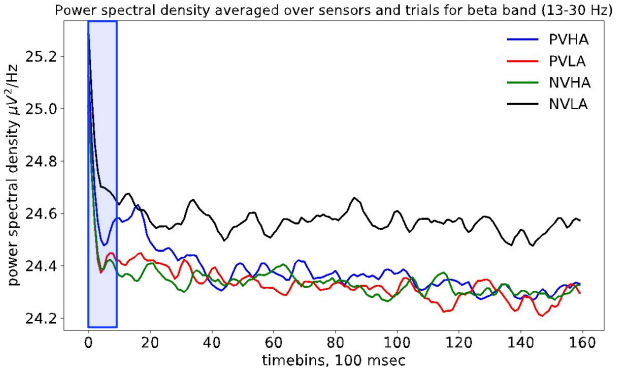
New model

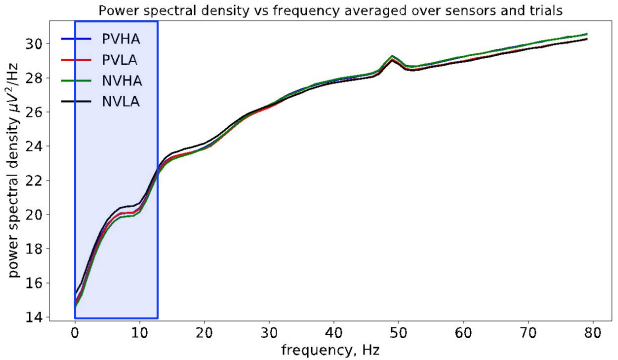


New model time trimmed



Trimming frequency to beta band (13Hz – 30Hz)





Old model without time trimmed:

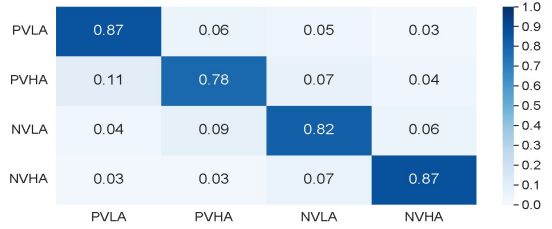
- Average Accuracy: 0.8341327072502654
- Average Precision: 0.8303159213115249
- Average Recall: 0.8333851884010003
- Average F1: 0.8318477236985471

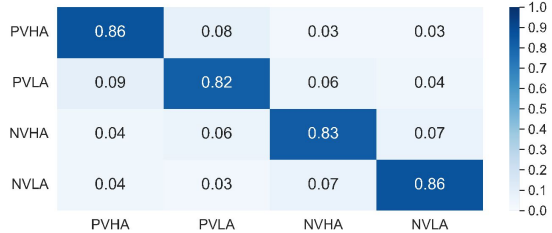
Old model with time/frequency trimmed :

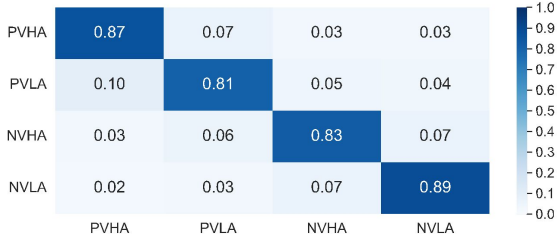
- Average Accuracy: 0.8431284588294481
- Average Precision: 0.8450086284110809
- Average Recall: 0.8429752796229250
- Average F1: 0.8420616334853701

New model with time/frequency trimmed :

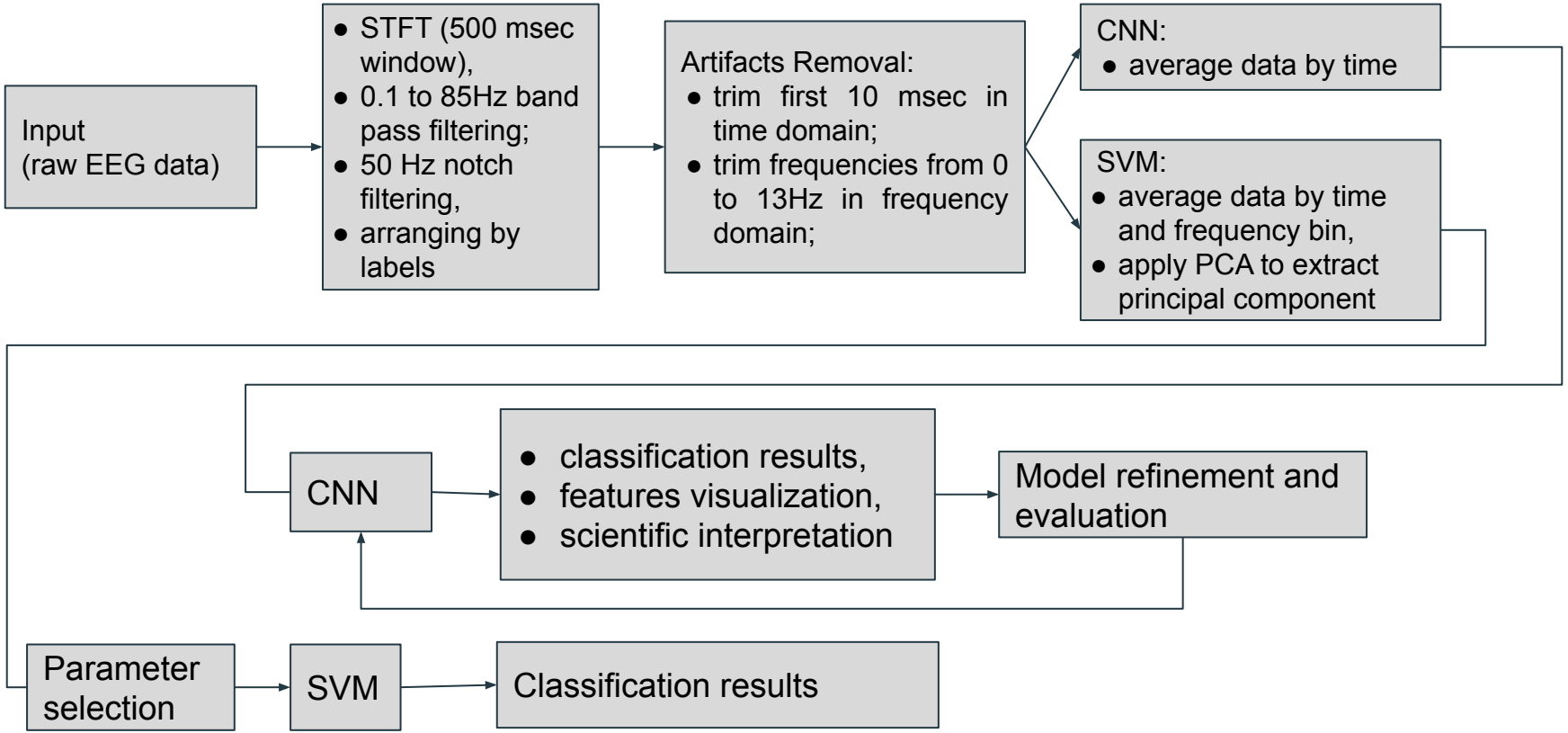
- Average Accuracy: 0.8494913075074069
- Average Precision: 0.8496308053813797
- Average Recall: 0.8497149452253278
- Average F1: 0.8482430103761518







Pipeline for Emotion Classification with CNN/SVM



Final Results

1. From the visualization of dense layer NVLA associates with lower frequency bands (beta)
2. PVLA and PVHA – with higher frequency bands,
3. NVHA distributed across frequency bands,
4. The feature maps learn the features associated with frequencies and channels,
5. Some feature maps represent activity over frequencies for specific channels (sharp vertical lines) others for wider range of channels (blurred vertical lines),
6. Removal of trigger-associated noise (first 10 timebins) improves accuracy,
7. Trimming timebins till the lower boundary of beta band (13Hz) along with removing first 10 timebins brings the accuracy of a new model (0.85) to the same of the SVM used in [1].
8. The tsne map suggests the presence of several emotional levels for each emotion category

[1]. V. Babushkin, W. Park, M. Hassan Jamil, H. Alsuradi, and M. Eid, “EEG-based classification of the intensity of emotional responses,” in *10th International IEEE EMBS Conference on Neural Engineering*, 2021.

Reference

[1]. V. Babushkin, W. Park, M. Hassan Jamil, H. Alsuradi, and M. Eid, “EEG-based classification of the intensity of emotional responses,” in *10th International IEEE EMBS Conference on Neural Engineering*, 2021.