

Looked into NMS for box combination method:

- <https://towardsdatascience.com/non-maximum-suppression-nms-93ce178e177c>

Mel spectrograms:

- <https://towardsdatascience.com/audio-deep-learning-made-simple-sound-classification-step-by-step-cebc936bbe5>
- <https://towardsdatascience.com/audio-deep-learning-made-simple-part-2-why-mel-spectrograms-perform-better-aad889a93505>

Convolutional Variational AutoEncoder (CVAE)

- <https://skannai.medium.com/what-are-convolutional-variational-auto-encoders-cvae-515f4fedc23>