Real-Time Sperm Identification from IVF Testicular Biopsy

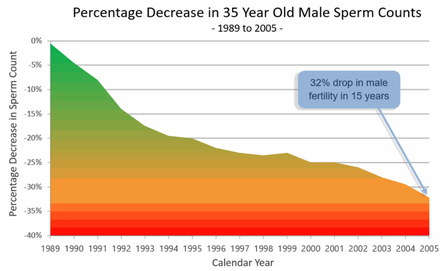
Daniel Wu Tayden Li

Stanford Computer Science

Background

Global average sperm count in men is falling at 1.6% a year. 30 million males suffer from male infertility, and are turning to in-vitro fertilization (IVF). Testicular biopsies are the primary sources of sperm for patients with severe infertility. However:

- Manual sperm image analysis is highly labor-intensive, time-consuming, and shows significant variability.
- Existing commercial computer assisted sperm analysis (CASA) systems are inaccurate, and perform poorly on low concentration sperm samples.

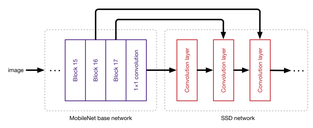


Objectives

We aim to automate manual sperm analysis during IVF, especially for patients with extreme infertility. We will develop low-latency object detection CNNs, in order to deploy real-time sperm identification and selection tools into the clinic.

Models

We trained five different models, which were pre-trained on the COCO object detection challenge.



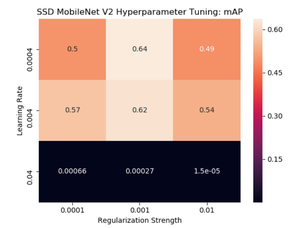
Methods

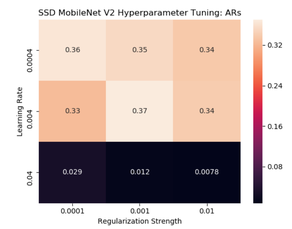
We built a custom data preprocessing and standardization pipeline, and evaluated the untuned performance of our models on the validation set. After finding poor initial performance with RetinaNet-MobileNet V1, RetinaNet-ResNet 50, and Faster R-CNN-ResNet 50, we conducted hyperparameter tuning on SSD MobileNet V2 and Faster RCNN Inception V2, exploring variations in learning rate, regularization strength, IOU threshold, image resolution, batch size, and the localization and classification loss ratio.
We then deployed our model to a video classification pipeline, and tested clinical integration.

IOU Threshold	0.4	0.6	0.8
mAP	0.574	0.576	0.517
AR	0.448	0.319	0.292

Component Loss	1	0.66	0.33
mAP	0.576	0.545	0.550
AR	0.319	0.352	0.2353

Tuning IOU threshold and component loss weight for SSD MobileNet V2.



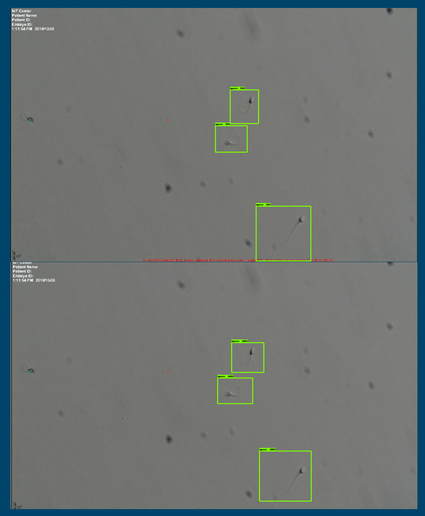


Results

Three embryologists independently labelled our test dataset of 110 images, containing 111 sperm. Because we have a small sample size per image, mAP and AR for human performance are overestimates.
We calculated mAP at 0.5 IOU and AR at 100 detections per image on our test set, with our best hyperparameters.
The model has a difficult time identifying distorted sperm, particularly those with bent tails, deformities, or those occluded by microscopy artifacts. Detection rates and localization performance on healthy sperm qualitatively align with embryologist-level performance.

Feature Extractor	Object Detector	mAP	AR
Embryologists	Occipital Cortex	0.925	0.642
MobileNet V2	SSD	0.741	0.376
Inception	Faster R-CNN	0.657	0.356

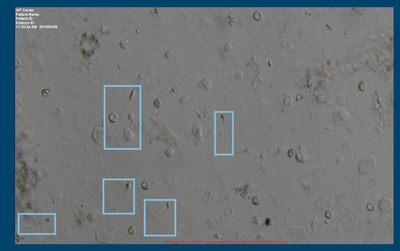
Comparison of performance of architectures on test set.



Example predictions from SSD MobileNet V2.
Top: Model Output, Bottom: Human Labels

Dataset

- We curated a novel testicular biopsy samples dataset of 702 images from 30 patients in collaboration with Stanford Reproductive Health.
- These images vary in sperm phenotype, cellular clutter, tissue superstructure, cell density, imaging modality, size, and resolution.
- Images are normalized and passed through glare filters and diffraction correction.
- Images are randomly flipped horizontally, cropped, padded, and jittered.



Sample testicular biopsy image with embryologist-labelled bounding boxes.

Conclusion

- We achieved an mAP@IOU=0.5 of 0.741, with an AR@100 detections per image of 0.376 with an SSD MobileNet V2.
- There is work to be done on improving model performance with hyperparameter tuning. We will add sperm morphology classification, and test and deploy models at Stanford Hospital.
- We have deployed the model to a real-time video classification pipeline, which automatically identifies sperm in testicular tissue.
- Our tool significantly lowers the monetary and labor costs of sperm selection in *in vitro* fertilization.