Quantifying DNA Damage in Comet Assay images using Neural Networks

Fast Machine Learning Workshop, Monday 30th November 2020

Selina Dhinsey

UNIVERSITY OF LIVERPOOL

NHS The Clatterbridge Cancer Centre NHS Foundation Trust

LIV.

nwcr NORTH WEST CANCER RESEARCH

Science & Technology Facilities Council

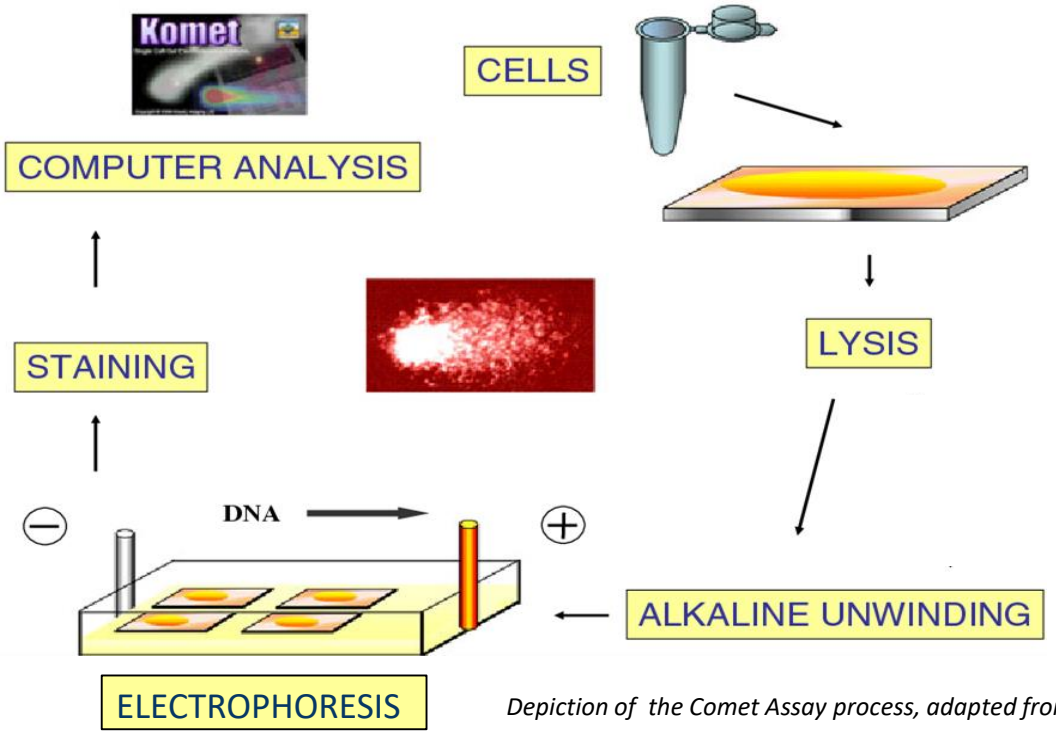
QUASAR GROUP

Overview

- Comet Assay
  - Measure DNA damage
- Program to automatically quantify damage
  - Incorporating NN for instance segmentation
  - Issues data available
- MC simulation of comet assay images

Comet Assay

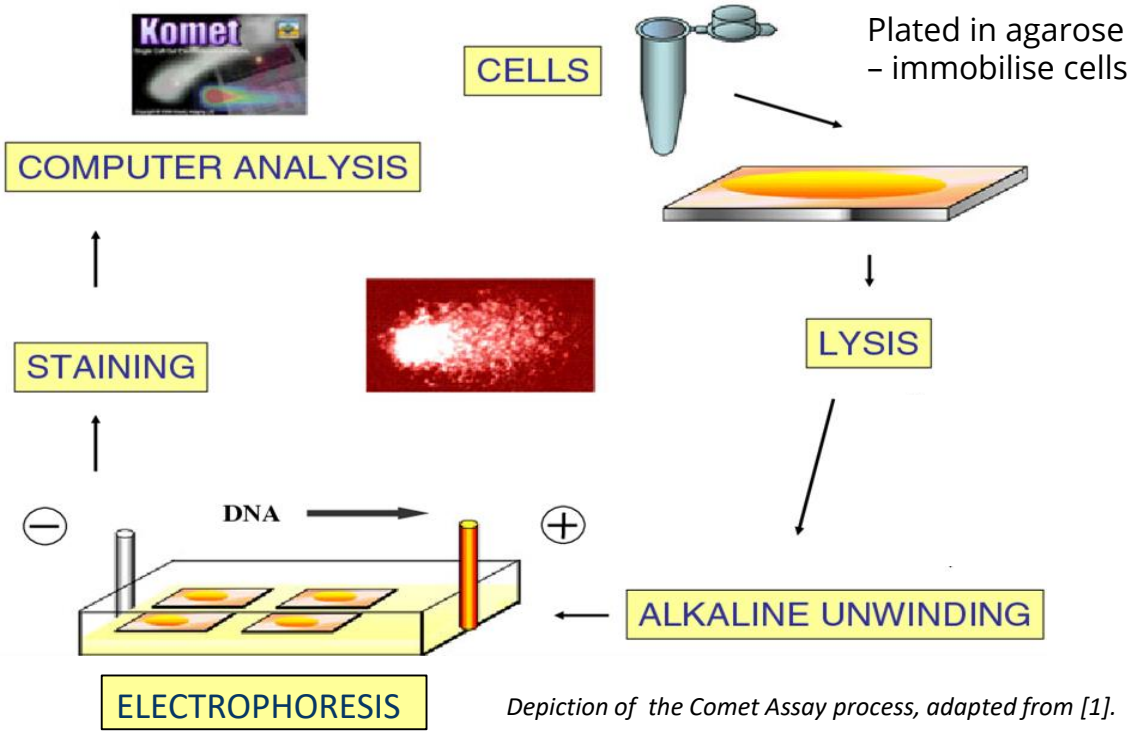
Procedure for obtaining
DNA damage levels



Depiction of the Comet Assay process, adapted from [1].

Comet Assay

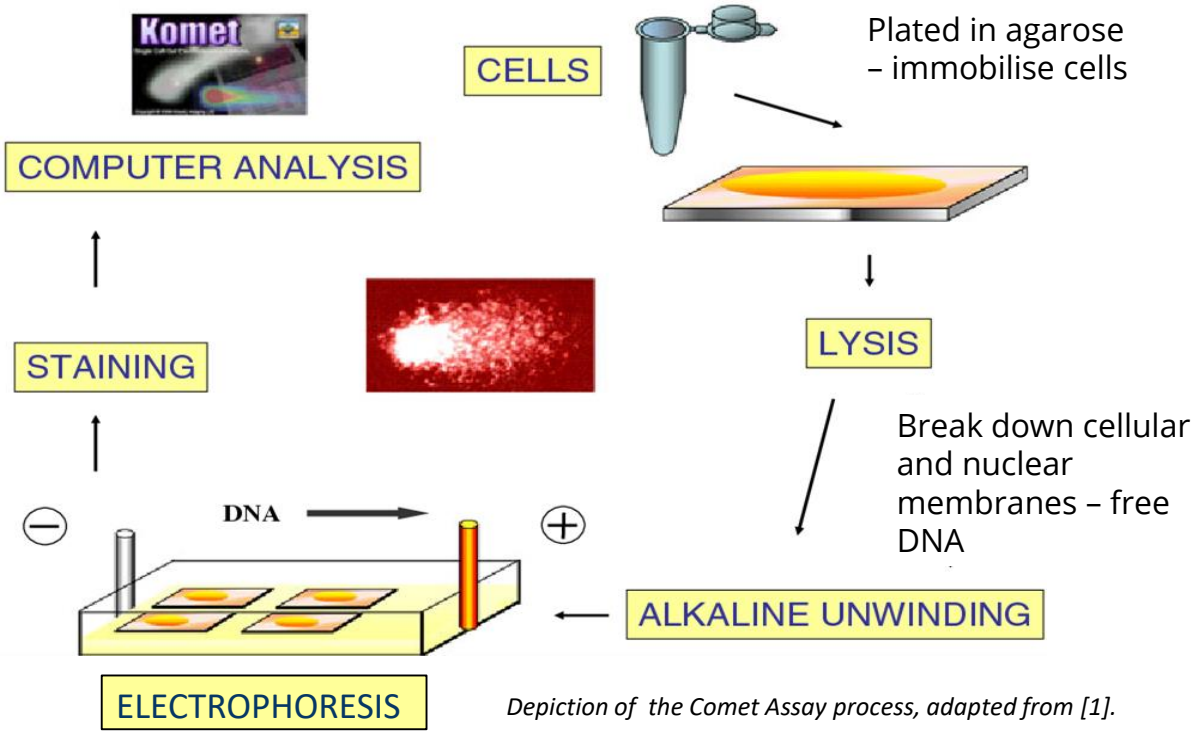
Procedure for obtaining
DNA damage levels



Depiction of the Comet Assay process, adapted from [1].

Comet Assay

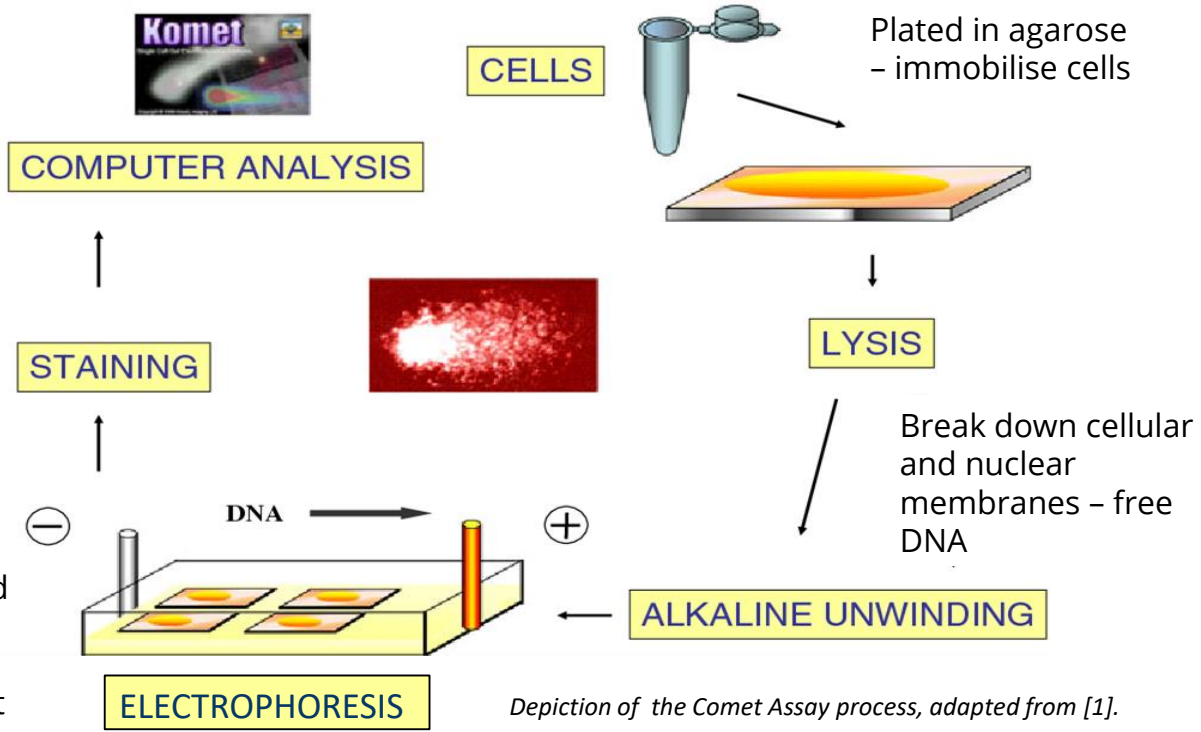
Procedure for obtaining
DNA damage levels



Depiction of the Comet Assay process, adapted from [1].

Comet Assay

Procedure for obtaining
DNA damage levels



Depiction of the Comet Assay process, adapted from [1].

Comet Assay Images

More fluorescent and elongated tail indicative of more strand breaks

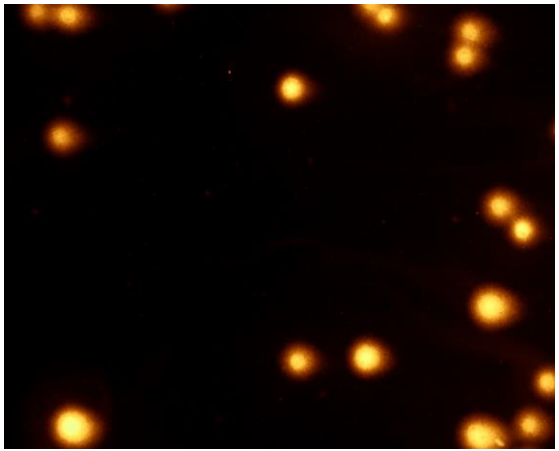


Image of control cells (no irradiation) after a comet assay

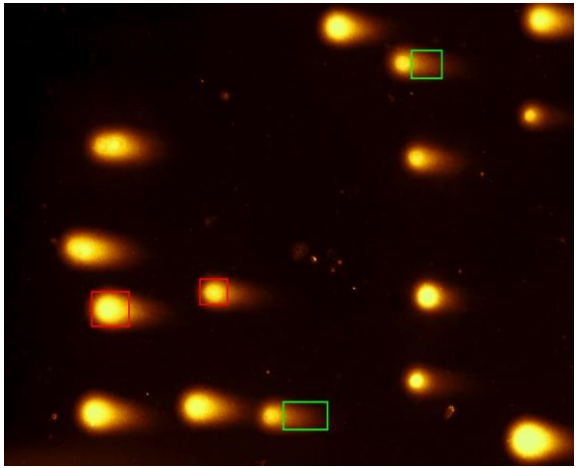


Image of irradiated cells that have formed comets following a comet assay. Comet heads are outlined in red, tails in green.

Comet Assay Images

More fluorescent and elongated tail indicative of more strand breaks

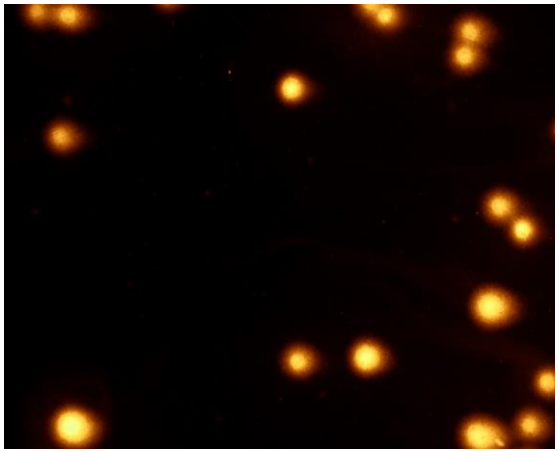


Image of control cells (no irradiation) after a comet assay

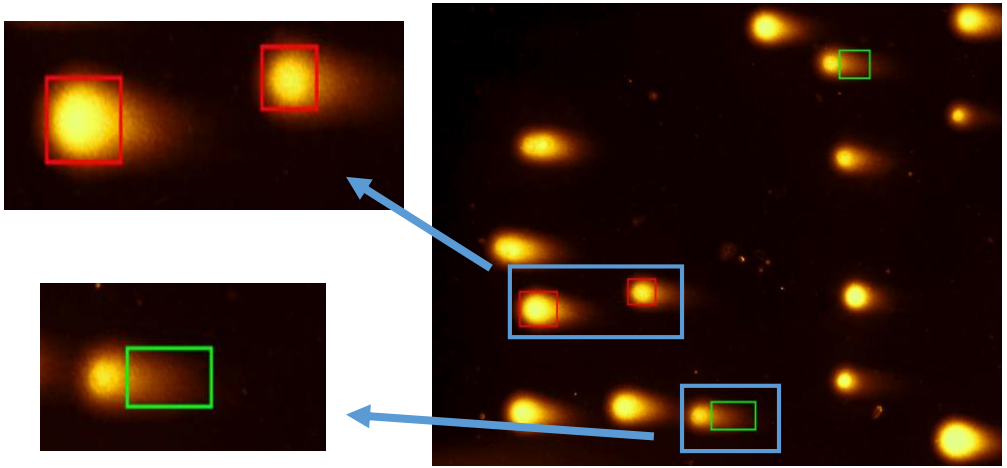
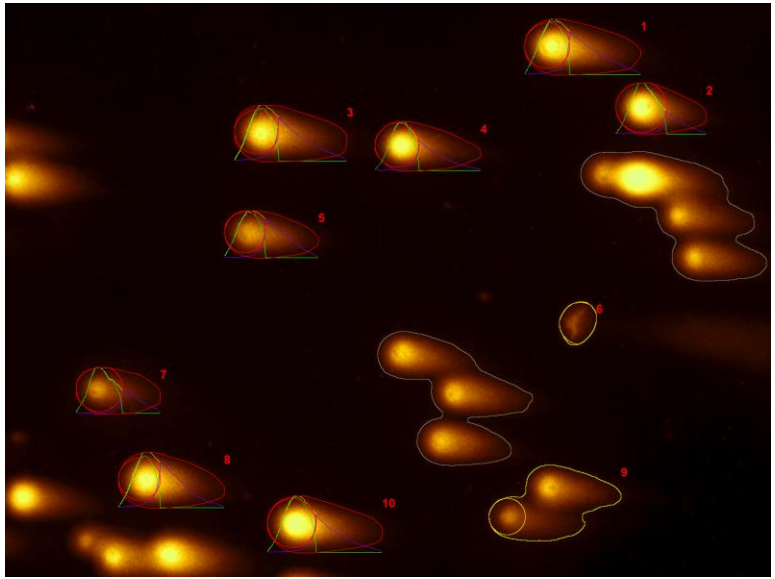


Image of irradiated cells that have formed comets following a comet assay. Comet heads are outlined in red, tails in green.

Image Analysis

- Methods of analysis vary a lot
  - By eye - relative & carries lots of variability - no numeric results
  - Semi-automated - user defines comet regions and comet properties are measured
  - Full Automation - quickest, but measurement algorithms still vary
- Developing programme to automate analysis
  - As little user input as possible
  - Written in Python

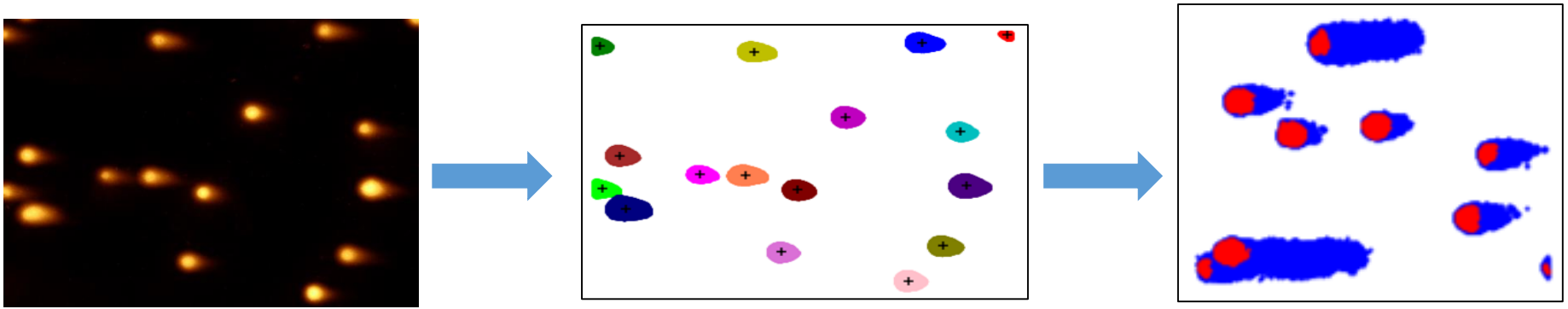


OpenComet
A fully automated free comet assay software

Open Source comet analysis software [1]

Measuring Comets

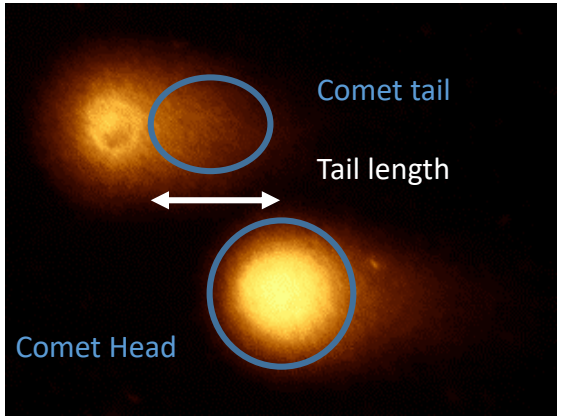
- Image analysis and thresholding to extract comet body and head
- Widely used measurements: **tail length and tail DNA %**
- Thresholding **loses data** around comet edge

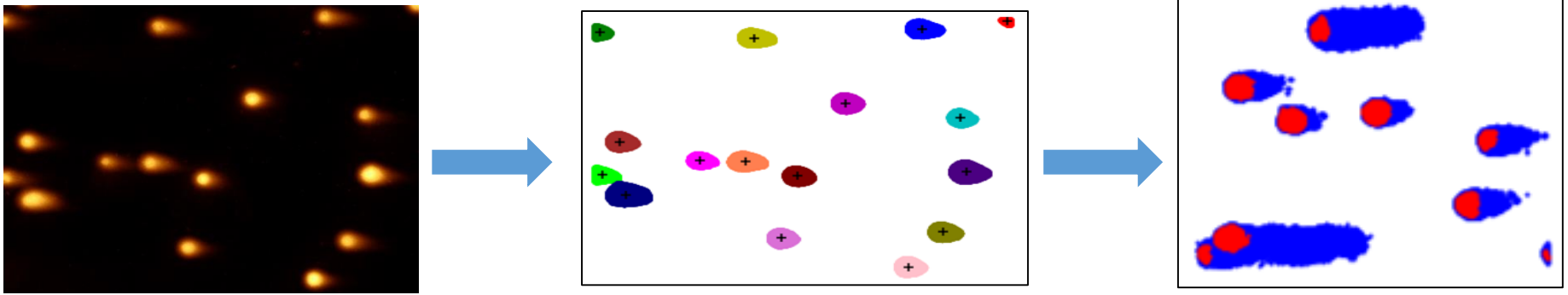


Measuring Comets

$$tail\ DNA = \frac{\sum tail\ pixel\ intensities}{\sum comet\ pixel\ intesnities} * 100$$

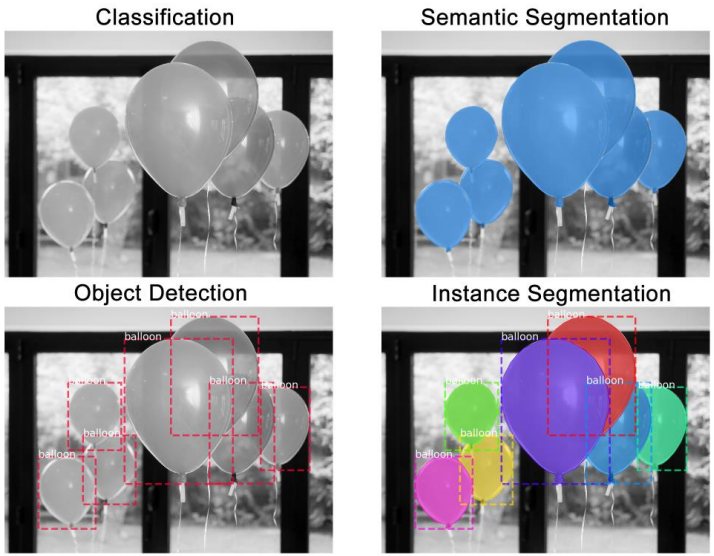
- Image analysis and thresholding to extract comet body and head
- Widely used measurements: **tail length and tail DNA %**
- Thresholding **loses data** around comet edge





Identifying Comets

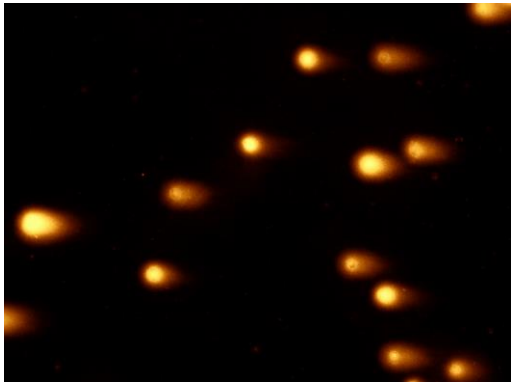
- Train a network to identify comets within images automatically as a first step
- Using **instance segmentation –** identify all instances of comets as separate masks
- Identifications based on overall structure rather than intensity - more robust across different image sets

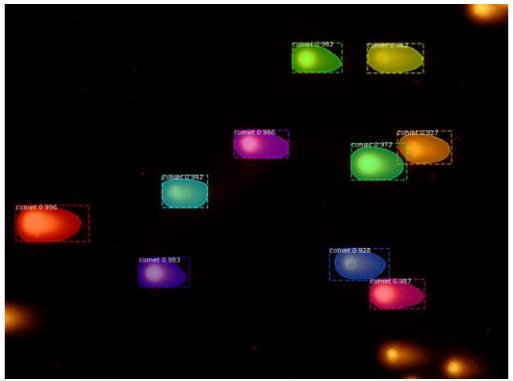


Representation of how the instance segmentation method combines the bounding box and pixel-wise classification methods [3].

Instance Segmentation

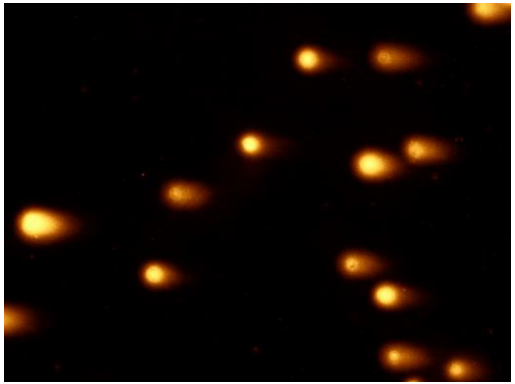
- Mask R-CNN combines Faster R-CNN and FCN
  - Faster R-CNN consists of RPN and RoIPool
  - RPN proposes bounding boxes
  - RoIPool extracts features to classify bboxes
- Adding FCN gives additional mask output
  - Total loss is combined from classifications, bounding boxes and masks

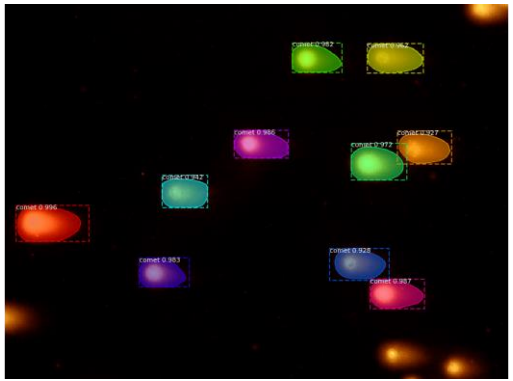




Instance Segmentation

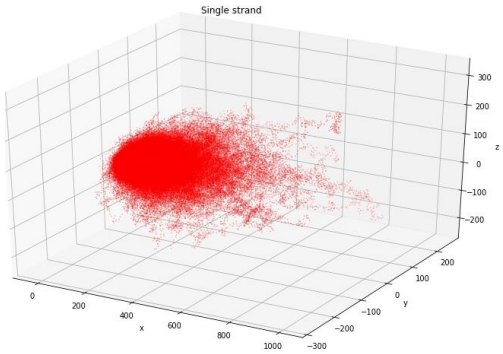
- Mask R-CNN combines Faster R-CNN and FCN [4]
  - Faster R-CNN consists of RPN and RoIPool
  - RPN proposes bounding boxes
  - RoIPool extracts features to classify bboxes
- Adding FCN gives additional mask output
  - Total loss is combined from classifications, bounding boxes and masks
- Transfer learning - use already trained weights (COCO model [5])
  - < 100 images used
  - Vastly quicker train time
  - >90% confidence in masks
- However, **lack of data** available for training/testing

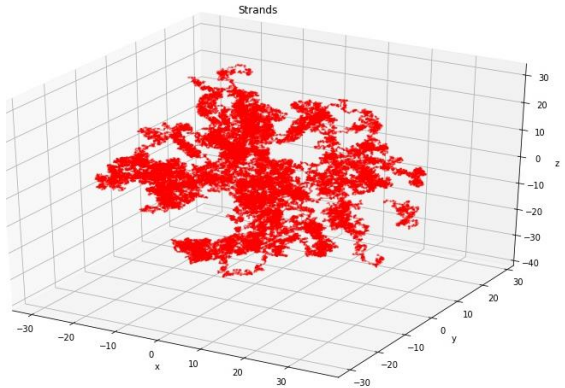


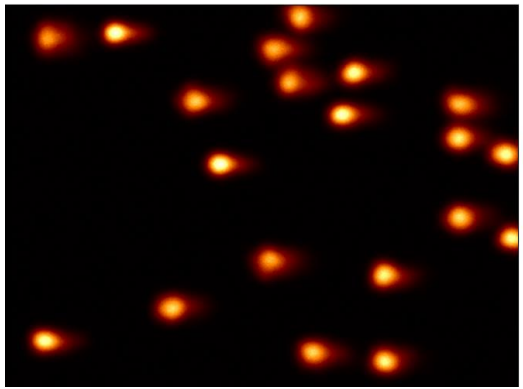


Simulating Images with MC

- Crude modelling of assay to produce simulated images
  - Random breaks on strands
  - Drift broken strands away from cell: E field strength and strand length dependent
- Use to further test neural net performance and train other models







Summary

- Comet assay images need better, consistent quantification of DNA damage levels
- Program proposed automates the process
  - Uses instance segmentation to identify comets
- Monte Carlo
  - Simulate the assay procedure to produce simulated images
  - Provide more data for NN models and further testing of performance

References

[1] H. Karlsson. “The comet assay in nanotoxicology research”. In:Analytical and Bioana-lytical Chemistry398 (2010), pp. 651–666
[2] BM Gyori, G Venkatachalam, PS Thiagarajan, D Hsu and MV Clement. "OpenComet: An automated tool for comet assay image analysis", Redox Biology, 2:457-465, 2014.
[3] https://engineering.matterport.com/splash-of-color-instance-segmentation-with-mask-r-cnn-and-tensorflow-7c761e238b46
[4] Kaiming He et al. “Mask R-CNN”. In:CoRRabs/1703.06870 (2017). arXiv:1703.06870.url:http://arxiv.org/abs/1703.06870
[5] Tsung-Yi Lin et al. “Microsoft COCO: Common Objects in Context”. In:CoRRabs/1405.0312(2014). arXiv:1405.0312.url:http://arxiv.org/abs/1405.0312

Thank you, Questions??