Automating Morphological Profiling with Generic Deep Convolutional Networks

Nick Pawlowski[1], Juan C Caicedo[2], Shantanu Singh[2], Anne E Carpenter[2], Amos Storkey[3]

(1) Imperial College (previously (3)), (2) Broad Institute, (3) University of Edinburgh

1: Motivation

Morphological profiling aims to create **profiles** or signatures of genes, chemicals and diseases from microscopy images. These signatures have several applications, including

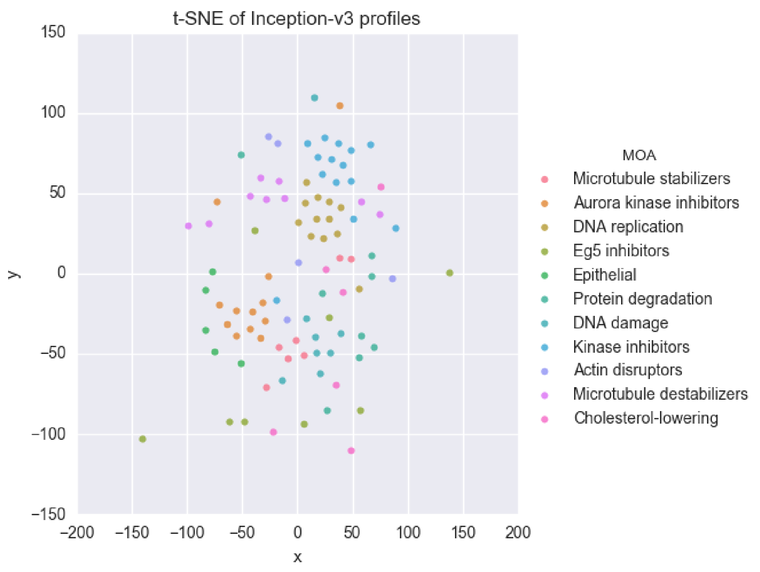
- functional genomics
- drug discovery
- target identification

Current approaches need human supervision to fine-tune parameters of classical algorithms to extract meaningful features from images **caicedo_profiling** We propose the use of pre-trained deep convolutional networks for feature extraction. Our contributions are:

- **Speed**: It enables faster profile computation than the classical pipeline with image segmentation and feature extraction.
- **Autonomous**: It eliminates the need for human input to tune parameters.
- **Performance**: The extracted profiles achieve a better accuracy than a baseline approach based on handcrafted features.

3: Representation Space

On an example dataset, the extracted profiles from neural networks span a representation space with mostly clear clusters regarding the biological mechanism-of-action. Inception-v3 extracted profiles can be visualised using t-SNE.

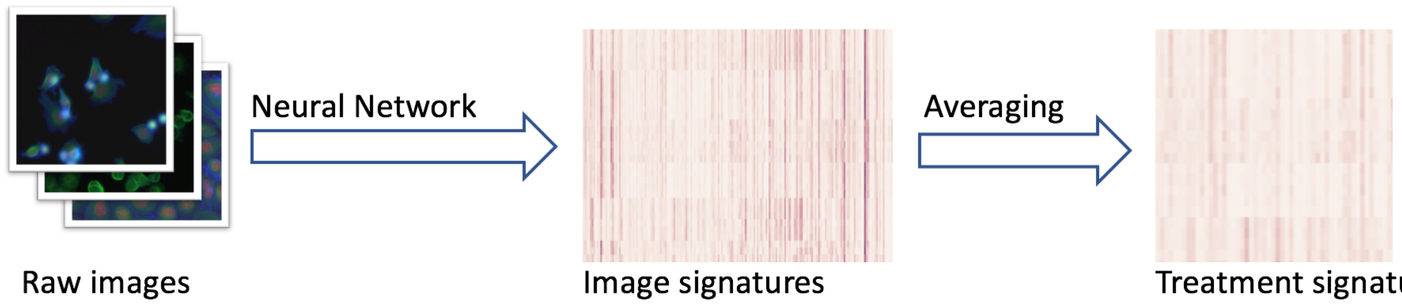


Inception-v3 profiles show the best clustering of biological mechanisms-of-action of the tested compounds.

References

2: Overview

Activations of pre-trained neural network constitute profiles for individual images. Profiles from images of the same treatment are averaged to form a treatment (i.e. gene, chemical, or disease) profile.



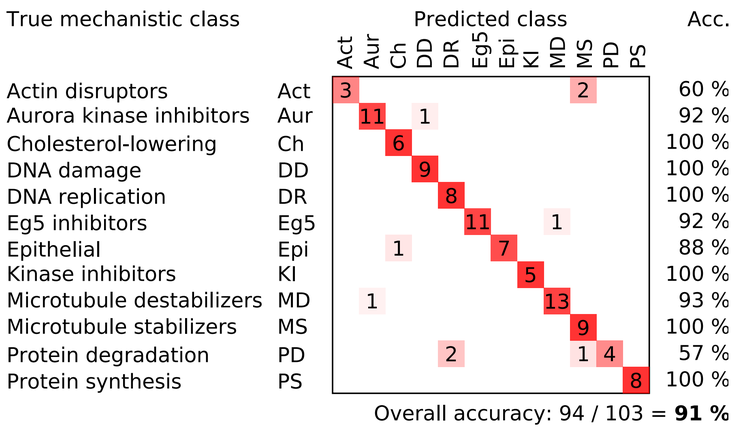
By measuring similarities or distances between these profiles in this representation space, they can be used to find relationships between the treatments – a valuable tool when investigating the function of a gene or allele, figuring out the mechanism of action of a compound, relating the phenotype of a disease with that of a genetic or chemical treatment, and many questions.

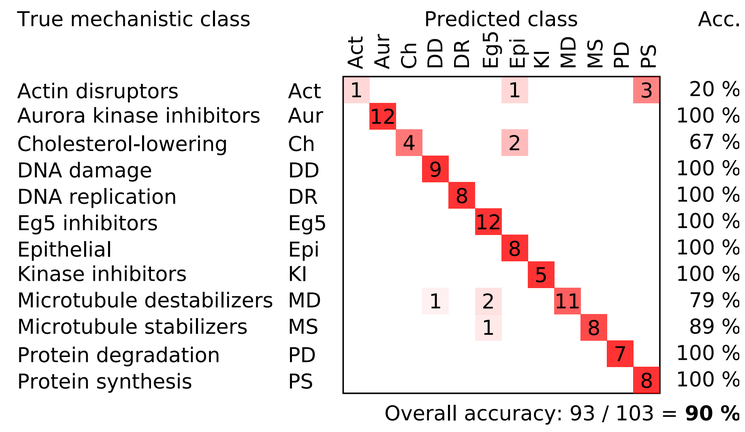
4: Mechanism-of-Action Classification

Classification results using different pre-trained neural networks with different types of images. The neural networks were pre-trained on the ImageNet dataset. The full-sized images were resized to the appropriate input size of the neural network. Greyscale images account for the use of each greyscale channel as separate image. Illum corrected images use the technique proposed by Singh et al.**Singh2014** to compensate for illumination bias during microscopy imaging. We use 1NN classification leaving out profiles from the same compound.

	Inception-v3**Szegedy2015**	ResNet-152**He2015**	ResNet-101**He20**
Full Images	70.87%	55.34%	57.2
+ illum corrected	69.90%	65.05%	65.0
+ greyscale	86.41%	75.72%	70.8
+ illum corrected & greyscale	91.26%	78.64%	79.6

Confusion matrices for the Inception-v3 extracted profiles (left) show an improvement over the classical results (right)**Singh2014**

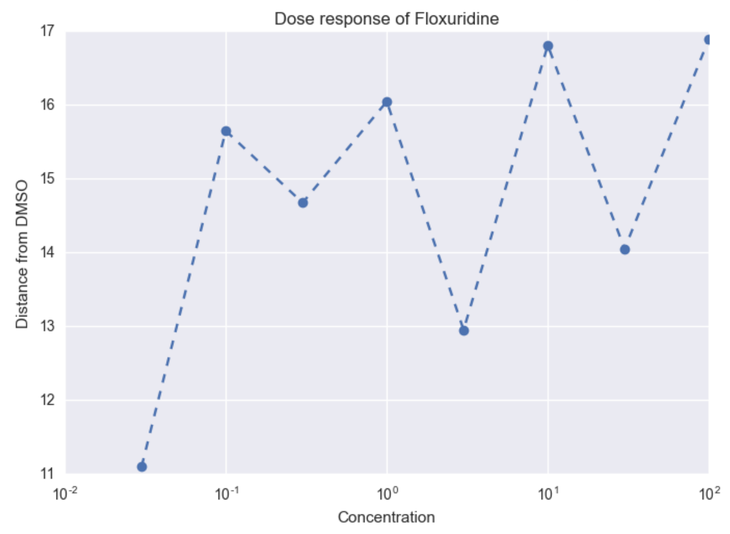


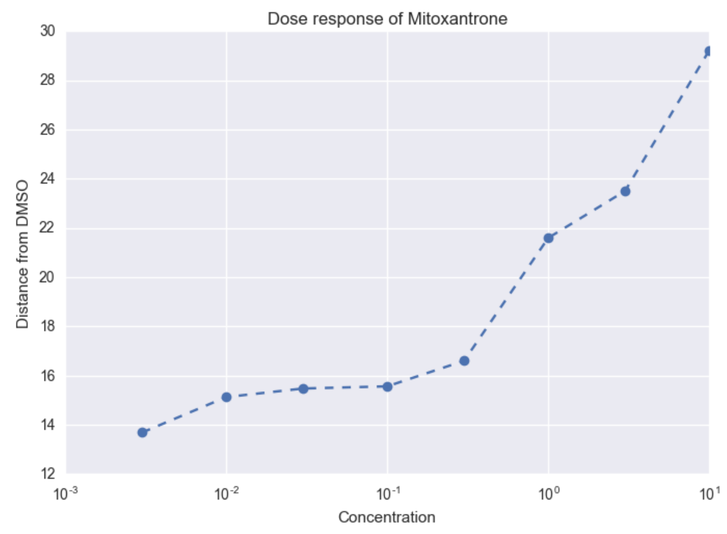


The different confusion matrices show that complimentary information is present in both profile types.

5: Dose Response

The extracted profiles can be evaluated according to the dose response curves. Some compounds show chaotic relations (left). However, other compounds exhibit the expected direct relation between higher dose and stronger response (right).





Code & Dataset

Code is available at `github.com/carpenterlab/2016_pawlowski_mlcb`.
Dataset is available as BBBC021 of the Broad Bioimage Benchmark Collection at `data.broadinstitute.org/bbbc/BBBC021/`

Acknowledgements

We want to thank Mike Ando (Google, Inc.) for helpful discussions about the use of transfer learning for the domain of morphological profiling. We want to thank all members of the Imaging Platform for helpful discussions. AEC