Synthetic Biology Lab Automation Augmented with AI:
deep learning for object detection and classification

Peter Hebden

phebden@gmail.com

1 February 2022

Outline

Introduction (1)

- This talk is about using lab automation + AI for picking carotene producing yeast colonies.
- Why automation + AI?
- Automated experiments are faster and more accurate, consistent, self-documenting and reproducible.
- Automation augmented with AI can make intelligent decisions at many points in the research workflow.
- Automation can generate large training datasets for the AI.

AI: Supervised Machine Learning

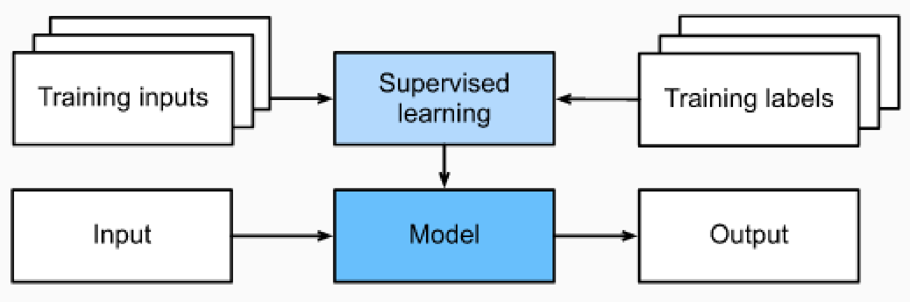
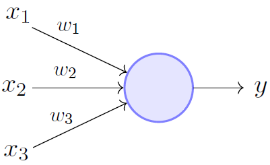


Figure: Learning from labelled examples: Input training data to the supervised learning algorithm and it outputs a trained model. Then inputs of new examples to this model will generate outputs in the form of predictions which may be numerical values or classes depending of the application. Although not new, neural networks can now outperform the common algorithms: linear classifiers, k-nearest neighbor, support vector machines, decision trees, and random forests.

AI: Neural Networks



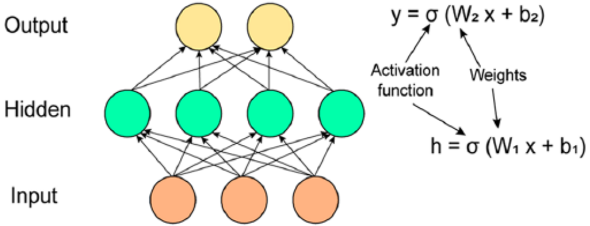


Image: [Volk et al., 2020].

- AI in the form of an artificial neural network was inspired by the how the brain works.
- Deep learning uses neural networks with many hidden layers and very large datasets to find hidden structure in the data and make accurate predictions [LeCun et al., 2015] (over 42360 citations).

AI: Deep Learning

- Deep neural networks are known for accurate classification of complex images [LeCun et al., 2015, Esteva et al., 2017].

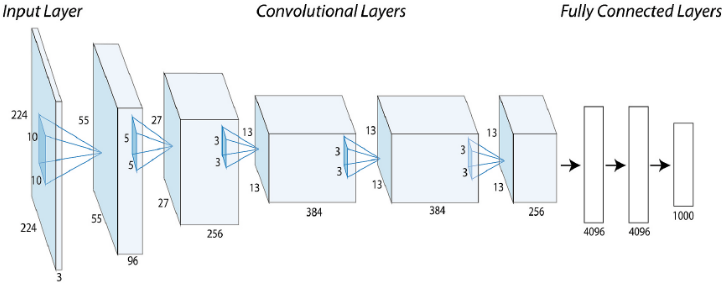


Figure: AlexNet used GPUs and network depth to improve performance (more layers are better), won ImageNet contest in 2012. The 5 convolutional layers use 96, 256, or 384 filters. Predicts 1000 classes [Krizhevsky et al., 2012] (over 94340 citations). Image: [Lawson et al., 2021]

AI: Optimization

- Optimization efforts should be limited because performance depends heavily on the size and quality of the dataset.
- The easiest way to boost prediction accuracy is to use a larger and/or better training set [Opgenorth et al., 2019].
- Performance improves with more data long after other methods plateau [Lawson et al., 2021].

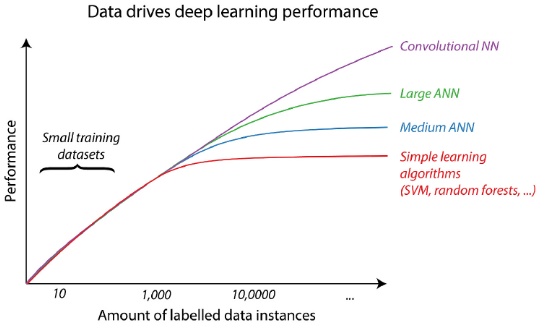


Figure: Deep learning performance continues to improve with the amount of training data when other methods plateau. Image: [Lawson et al., 2021]

Low cost open source colony picker

Figure: Raspberry Pi 4 Model B, 8GB RAM. Starter kit with power supply, micro SD card, cable and case. Cost £98.

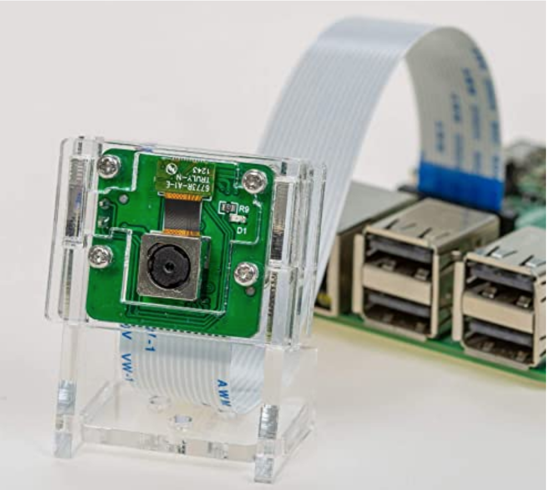


Figure: Arducam Auto Focus Camera with Motorized Focus Lens, OV5647 5MP 1080P. Cost £12.

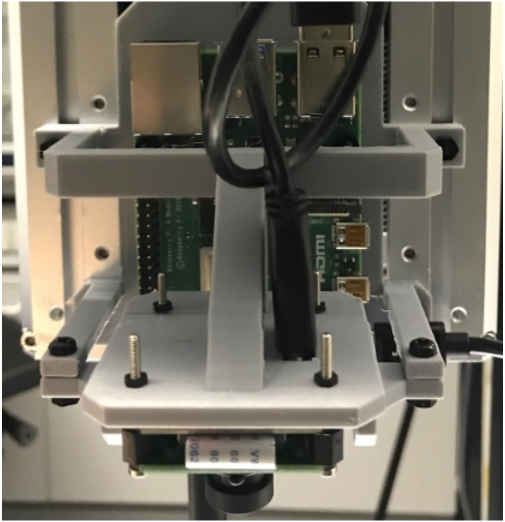


Figure: Raspberry Pi and camera attached to the back of the Opentrons OT-2 robotic arm with 3D printed brackets. Image: iGEM Marburg 2019 with expensive camera.

The Colony Picker

Figure: The colony picker configuration: The laptop communicates directly with the R Pi and the robot, and relays new colony coordinates from the R Pi to the robot. Image : Opentrons.com.

Workflow

- Use the colony picker to take pictures of colony plates that result from various DNA inputs under controlled conditions.
- Use LabelImg software to expertly label the images in a very consistent way, e.g. the orange or white colonies.
- Augment and partition images into training, validation, and test sets.
- Train the AI on the training set images, e.g using TensorFlow model_maker `https://www.tensorflow.org/lite/guide/model_maker`.
- Test the AI on unseen test set images.
- Run the colony picker on new colony plates.

The Colony Picker

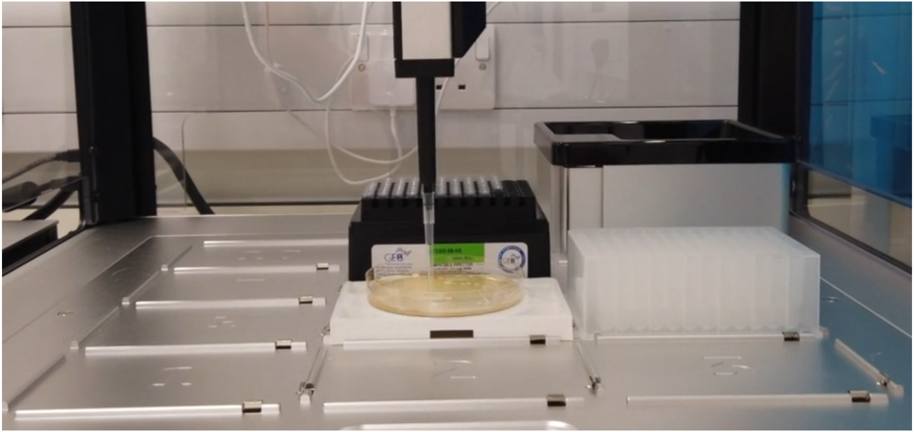


Figure: The Opentrons OT-2 liquid handling robot with colony picker in action. The robot has just received a list of colony coordinates and is picking a colony. Video: `https://www.youtube.com/channel/UCkWYMoMaR-2BUtU9O6clCAA`

Camera software: Picamera API v1.13.
Object detection and classification software: TensorFlow Lite 2.5.

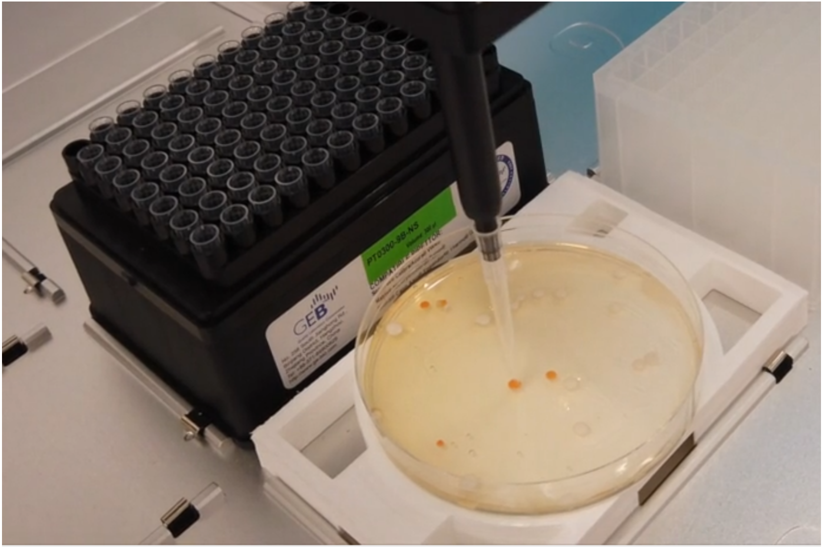
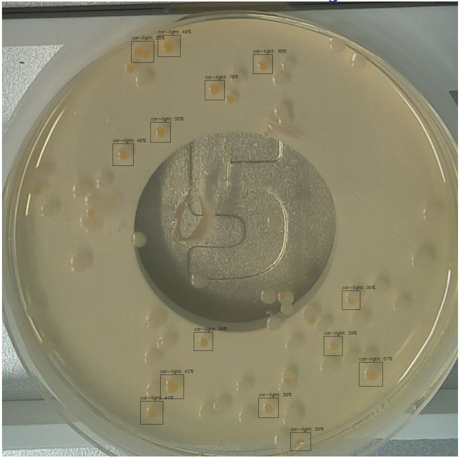
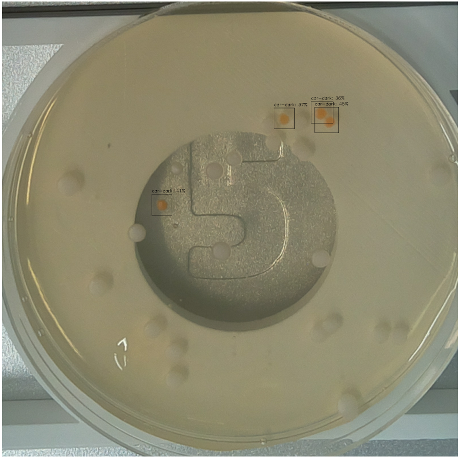


Figure: Closeup shot of the colony picker picking an orange colony with the OT-2 deck in the background.

The R Pi runs the object detection code



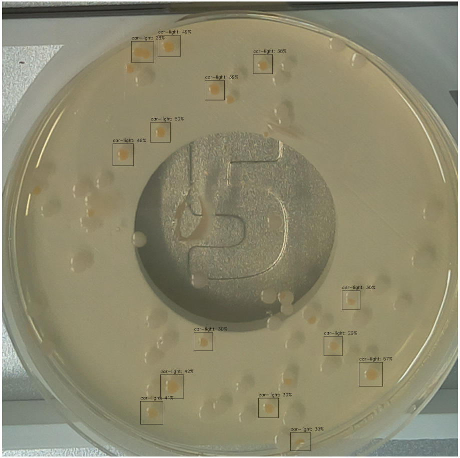
Light orange colonies.



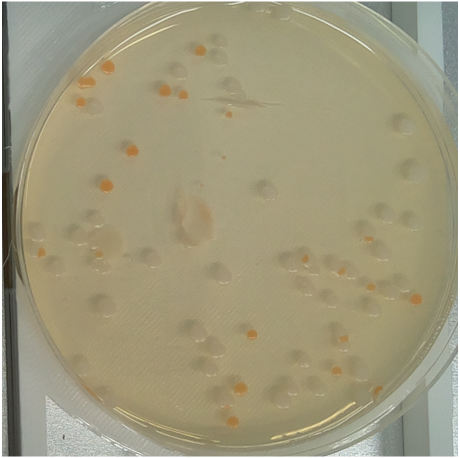
Dark orange colonies.

Training: 200 images (original + augmented) for 100 epochs. Test set: 8 images. Orange colonies were detected and classified as either light or dark, confidence threshold $>=$ 20%. White colonies were not labelled.

Classification based on color versus artifacts



Light orange colonies.



Same colony plate.

Light orange colonies darken after more time in the fridge.

Object detection using EfficientDet Lite1 model

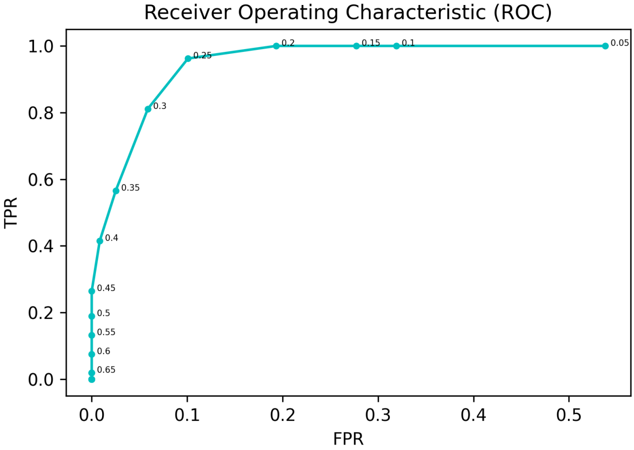
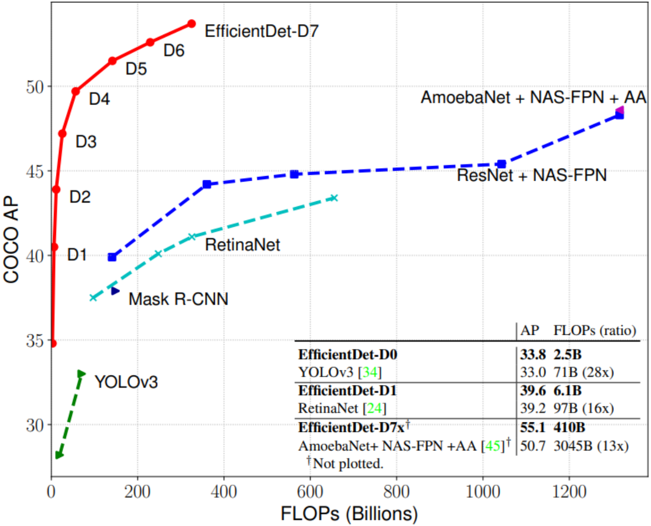


Figure: ROC curve: True Positive Rate (TPR) versus False Positive Rate (FPR). Colonies were detected and classified as carotene colonies using a range of confidence thresholds. The actual FPR was lower than indicated by the ROC curve because some orange colonies in the test set were not labelled by the expert.

EfficientDet models D0-D7



	AP	FLOPs (ratio)
EfficientDet-D0	**33.8**	**2.5B**
YOLOv3 [34]	33.0	71B (28x)
EfficientDet-D1	**39.6**	**6.1B**
RetinaNet [24]	39.2	97B (16x)
EfficientDet-D7x†	**55.1**	**410B**
AmoebaNet+ NAS-FPN +AA [45]†	50.7	3045B (13x)

†Not plotted.

Figure: Model configurations range from D0 to D7, each provides a different speed versus accuracy trade-off. Image: [Tan et al., 2020]

EfficientDet Lite models

The performance of EfficientDet-Lite models.

Model architecture	Size(MB)*	Latency(ms)**	Average Precision***
EfficientDet-Lite0	4.4	37	25.69%
EfficientDet-Lite1	5.8	49	30.55%
EfficientDet-Lite2	7.2	69	33.97%
EfficientDet-Lite3	11.4	116	37.70%
EfficientDet-Lite4	19.9	260	41.96%

Figure: EfficientDet Lite models (0..4) can run on the R Pi, smartphones, and PCs. More complex models are slower but can be more accurate [Tan et al., 2020]. Image: `tensorflow.org`

Note: Images are reduced to small square images before object detection: 320x320, 384x384, 448x448, 512x512, or 640x640 (pixels).

Conclusions

- A low cost open source colony picker extension to the Opentrons robot can be built for about £100.
- Object detection and classification do not require large datasets when the images are simple and the number of classes is small.
- Even 50 original images might be enough, especially when augmented.
- The TensorFlow machine learning library was sufficient for the colony picker.
- The EfficientDet Lite1 model was very accurate and more complex models could be used for more complex object detection and classification tasks.

References I

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S. (2017).
Dermatologist-level classification of skin cancer with deep neural networks.
Nature, 542(7639).

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012).
ImageNet classification with deep convolutional neural networks.
In *Advances in Neural Information Processing Systems*.

Lawson, C. E., Martí, J. M., Radivojevic, T., Jonnalagadda, S. V. R., Gentz, R., Hillson, N. J., Peisert, S., Kim, J., Simmons, B. A., Petzold, C. J., Singer, S. W., Mukhopadhyay, A., Tanjore, D., Dunn, J. G., and Garcia Martin, H. (2021).
Machine learning for metabolic engineering: A review.
Metabolic Engineering, 63(October 2020):34–60.

LeCun, Y., Bengio, Y., and Hinton, G. (2015).
Deep learning.
Nature, 521(7553).

Opgenorth, P., Costello, Z., Okada, T., Goyal, G., Chen, Y., Gin, J., Benites, V., de Raad, M., Northen, T. R., Deng, K., Deutsch, S., Baidoo, E. E. K., Petzold, C. J., Hillson, N. J., Garcia Martin, H., and Beller, H. R. (2019).
Lessons from Two Design–Build–Test–Learn Cycles of Dodecanol Production in <i>Escherichia coli</i> Aided by Machine Learning.
ACS Synthetic Biology, 8(6).

References II

Tan, M., Pang, R., and Le, Q. V. (2020).
EfficientDet: Scalable and Efficient Object Detection.

Volk, M. J., Lourentzou, I., Mishra, S., Vo, L. T., Zhai, C., and Zhao, H. (2020).
Biosystems Design by Machine Learning.
ACS Synthetic Biology, 9(7).