

รายงาน วิชา 261499 Deep learning

เรื่อง

Recognize the coffee cup in the image

จัดทำโดย

นาย ศุภากร คงกระจ่าง รหัส 600610784

นำเสนอ

อ.ดร.เกษมสิทธิ์ ดิยพันธ์

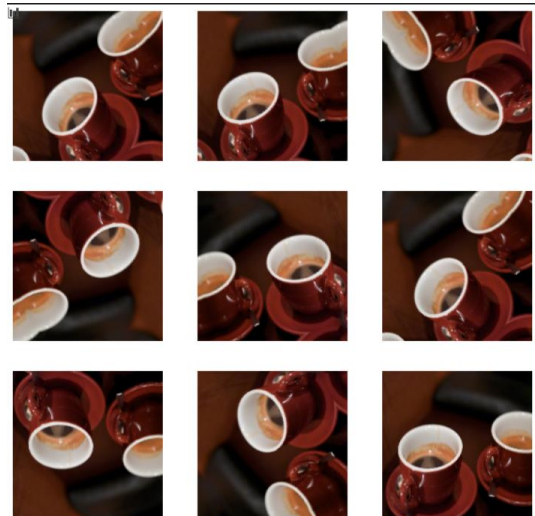
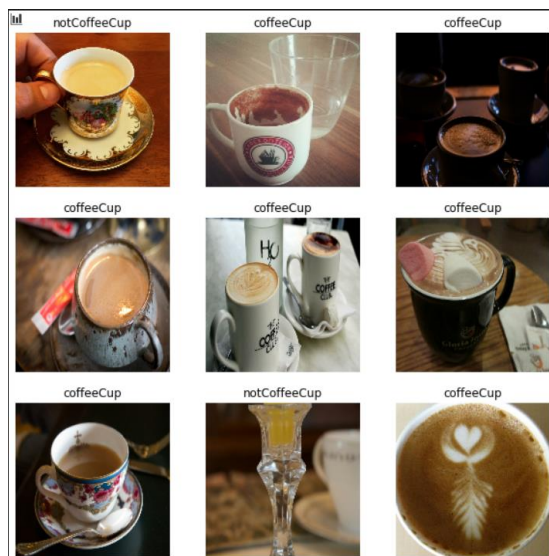
ที่มาของปัญหา

ที่มาของ การตรวจจับภาพว่า ในภาพมีถ้วยกาแฟอยู่หรือไม่ มาจาก ปัจจุบัน ผู้คนในหมู่บ้านทำงาน มีแนวโน้มที่จะมีการดื่มกาแฟเป็นประจำเพิ่มมากขึ้น ทำให้นักศึกษา ที่มีความชอบในกาแฟ มีความสนใจที่จะตรวจจับถ้วยกาแฟ เพื่อการเรียนรู้ในวิชา deep learning และ สามารถเป็นแนวทางการต่อยอดไอเดียใหม่ได้ เช่น เมื่อ ผู้ใช้ social media ตาม platform ต่าง ๆ ลงรูปภาพ ที่มีถ้วยกาแฟอยู่ในภาพ จะสามารถวิเคราะห์ความสนใจ ในการกาแฟของผู้ใช้งานได้ และ สามารถนำข้อมูลนั้น ไปทำการตลาดต่อไป

รายละเอียดวิธีการทำงาน

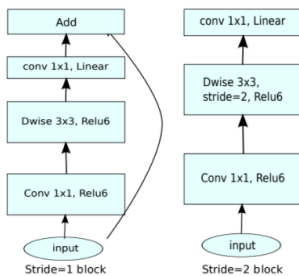
ขั้นตอน preprocessing

- มีการรวบรวมdataset รูปภาพถ้วยกาแฟ จาก Kaggle ซึ่งมีทั้งหมด 106 รูป โดยนำมาจัดเรียงหมวดหมู่ใหม่ ประกอบไปด้วย 2 คลาส คือ coffee cup กับ not coffee cup โดย คลาส coffee cup มี ทั้งหมด 86 รูป และ คลาส not coffee cup มี 20 รูป
- มีการแบ่ง training set เป็น 80% ของรูปทั้งหมด และ validation set เป็น 20%
- มีการแบ่ง batch size ในการ train แต่ละครั้ง คือ 32 และ image size เท่า 160 * 160
- มีการ prefetch dataset ให้ load data ในการ train ได้เร็วขึ้น
- มี การทำ data augmentation โดยมีการ random flip image ทั้งแบบ แนวตั้งและแนวนอน และมีการสุ่มหมุนรูปภาพจากระยะ 20% ของภาพเดิม
- มีการ map ค่า data จาก 0 – 255 เป็น -1 ถึง 1 เพื่อเข้า model



ขั้นตอนการสร้าง model

- ทำการ Transfer learning แบบ feature extraction คือ นำ pretrain model มาต่อกับ model ของตัวเอง โดย freeze weight ที่ pretrain model train มาแล้ว สามารถ update weight ได้เฉพาะ weight ของ model ตัวเอง ที่ต่อเข้าไป
- ใช้ pretrain model เป็น MobileNetV2 เนื่องจาก เป็น model ที่เล็ก สามารถ predict ได้รวดเร็ว เมื่อเทียบกับ model อื่น เช่น ResNet

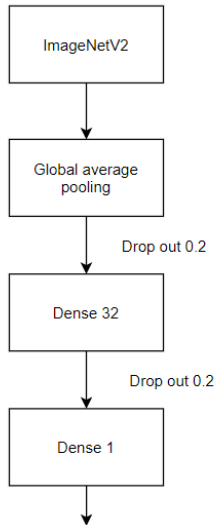


(d) Mobilenet V2

- ต่อด้วย global average pooling layer เพื่อ ทำ class activation map ได้ดี กล่าวคือ สามารถช่วยให้ ภาพ กลาย เป็น 1 มิติ เพื่อเข้า fully connected layer ต่อไป โดยสามารถ ลด parameters ในการ train ลงได้ค่อนข้างมาก และ สูญเสีย รายละเอียดของภาพไปน้อยลง ทำให้เรายังสามารถระบุ ตำแหน่งของ class ที่ต้องการตามภาพได้
- ทำ drop out ข้อมูล 20% ก่อนส่งไป layer ถัดไป เพื่อลด overfit
- ทำการทดสอบ เพิ่ม dense layer 2 แบบ แบบแรกคือ ทำ dense แบบ 32 node และ แบบที่สอง ทำ dense แบบ 2 node เพิ่มทดสอบ accuracy ว่าแบบไหน ได้ accuracy มากกว่ากัน
- ทำ drop out ข้อมูล 20% ก่อนส่งไป layer ถัดไป เพื่อลด overfit

- ทำ dense layer แบบ 1 node เพื่อ predict ภาพ ว่า ในภาพ มีถ้วยกาแฟหรือไม่

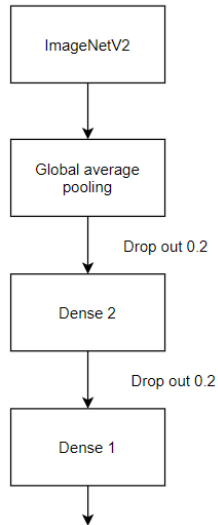
โครงสร้างในการทดลอง แบบที่ 1



โครงสร้าง แบบที่ 1

Layer (type)	Output Shape	Param #
input_8 (InputLayer)	[(None, 160, 160, 3)]	0
sequential_3 (Sequential)	(None, 160, 160, 3)	0
tf.math.truediv_3 (TFOpLambd	(None, 160, 160, 3)	0
tf.math.subtract_3 (TFOpLamb	(None, 160, 160, 3)	0
mobilenetv2_1.00_160 (Func	(None, 5, 5, 1280)	2257984
global_average_pooling2d_3 ((None, 1280)	0
dropout_6 (Dropout)	(None, 1280)	0
dense_6 (Dense)	(None, 32)	40992
dropout_7 (Dropout)	(None, 32)	0
dense_7 (Dense)	(None, 1)	33
=====		
Total params: 2,299,009		
Trainable params: 41,025		
Non-trainable params: 2,257,984		

โครงสร้างในการทดลอง แบบที่ 2

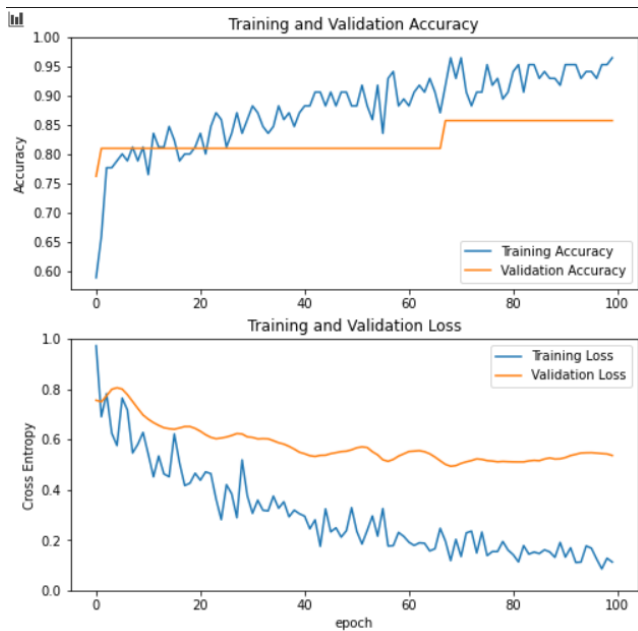


โครงสร้างแบบที่ 2

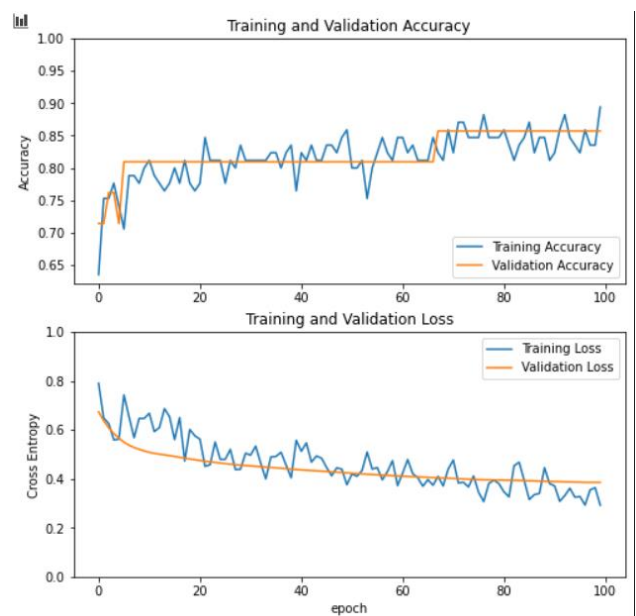
Layer (type)	Output Shape	Param #
input_4 (InputLayer)	[(None, 160, 160, 3)]	0
sequential_1 (Sequential)	(None, 160, 160, 3)	0
tf.math.truediv_1 (TFOpLambd	(None, 160, 160, 3)	0
tf.math.subtract_1 (TFOpLamb	(None, 160, 160, 3)	0
mobilenetv2_1.00_160 (Func	(None, 5, 5, 1280)	2257984
global_average_pooling2d_1 ((None, 1280)	0
dropout_2 (Dropout)	(None, 1280)	0
dense_2 (Dense)	(None, 2)	2562
dropout_3 (Dropout)	(None, 2)	0
dense_3 (Dense)	(None, 1)	3
=====		
Total params: 2,260,549		
Trainable params: 2,565		
Non-trainable params: 2,257,984		

- ใช้ learning rate = 0.0001 และ epoch = 100
- ใช้ optimizer แบบ adam เพื่อ สามารถปรับ momentum ตาม epoch ได้ตามสูตรของ adam ทำให้ train มีความ smooth มากขึ้น
- ใช้ loss function เป็น binary cross entropy สำหรับวัด ค่า error แบบ สองคลาส แค่ 0 กับ 1
- ในการทดลองมีเพียง training set กับ validation set ไม่ได้มี test set แต่อย่างใด

ผลการทดลอง และ การสรุปผล



กราฟโครงสร้างแบบที่ 1 (dense 32)



กราฟโครงสร้างแบบที่ 2 (dense 2)

จากกราฟ 1 ค่า accuracy = 0.9647 , val_accuracy = 0.8571, loss = 0.1127, val_loss = 0.5364

จากกราฟ 2 ค่า accuracy = 0.9647 , val_accuracy = 0.8571, loss = 0.1127, val_loss = 0.5364

จากผลการทดลอง สังเกตได้ว่า เมื่อมีจำนวน node ใน dense layer มากขึ้น ค่า accuracy ตอน train มีค่า มากกว่า accuracy ตอน validation จึงสรุปว่า ยิ่งเพิ่ม node ใน dense layer มาก ทำให้เกิดการ overfit ขึ้น หมายความว่า มี variance มากเกินไป โดยมีวิธีการแก้ไขได้สองวิธี คือ 1 ลดความซับซ้อนของโครงสร้างลด 2 เพิ่มจำนวน dataset ให้มากขึ้น โดยผู้ทดลองมีความเห็นว่า ควร เพิ่มจำนวน dataset ให้มากขึ้น เนื่องจาก dataset ที่มีนั้น มีค่อนข้างน้อย การเพิ่ม จำนวน dataset จะสามารถทำให้ ลดการ overfit ลงได้

สามารถดูโปรแกรมได้ที่: <https://github.com/werterzz/CoffeeCupProject.git>