

# Deep learning-based medical device classification

Tanut Bumrungvongsiri 63340500026

Phufa Boonchuatrong 63340500048

## 1. Problem Statement and Introduction

บริษัท บางกอก ยูนิเทรด จำกัด เป็นบริษัทที่คอยจัดเตรียมอุปกรณ์ทางการแพทย์ผ่าตัดให้กับโรงพยาบาล ซึ่งในการผ่าตัดในแต่ละเคสนั้น ทางโรงพยาบาลจะทำการเตรียมรายการของอุปกรณ์ใช้ในการผ่าตัดต่างๆ ส่งไปให้กับทางบริษัทจัดเตรียมอุปกรณ์นั้นตามรายการ ซึ่งการจัดส่งอุปกรณ์ผ่าตัดที่ผิดชนิดหรืออุปกรณ์ขาดหายนั้น จะส่งผลโดยกระทบที่ร้ายแรงต่อการผู้ป่วยได้ ซึ่งกระบวนการจัดเตรียมนั้นส่วนมากใช้มนุษย์ในการจัดเตรียม อีกทั้งตัวของอุปกรณ์ผ่าตัดที่บางชนิดมีลักษณะที่คล้ายกันมาก ซึ่งอาจจะทำให้เกิดความผิดพลาดในการจัดเตรียมได้ ทางบริษัทจึงเล็งเห็นในส่วนนี้ใช้ตัวของ AI เทคโนโลยีมาช่วยในการตรวจสอบความถูกต้องของการจัดเตรียมอุปกรณ์ก่อนส่งไปที่โรงพยาบาล

วัตถุประสงค์ของโครงการนี้คือการใช้เทคนิคทาง Computer Vision เพื่อใช้ในการจำแนกชนิดของเครื่องมือผ่าตัดโดยการใช้กล้องทำการบันทึกภาพของของเครื่องมือผ่าตัดที่ผ่านการจัดเตรียมมาแล้วที่พร้อมจะส่งไปโรงพยาบาล เพื่อตรวจสอบความถูกต้องของจำนวนและชนิดของเครื่องมือ ก่อนที่จะส่งไปโรงพยาบาล เพื่อลดความผิดพลาดในการจัดเตรียมแบบทั่วไปและลดความเสี่ยงที่ตัวของเครื่องมือผ่าตัดขาดหายหรือเครื่องมือผิดชนิดในการผ่าตัด

## 2. Background Study

### 2.1. Image Classification

เป็นการจำแนกชนิดวัตถุในรูปภาพด้วยเทคนิคการใช้ Deep Convolutional Neural Network มาใช้ในการสกัด Feature ของรูปภาพเพื่อให้โมเดลได้เรียนรู้รูปร่างต่างๆ ในรูปภาพซึ่งตัวของ image classification โมเดล ก็มี State of the art คือโมเดลที่ใช้การยอมรับกันโดยทั่วไปว่าเป็นโมเดลมีประสิทธิภาพที่ดี ตัวอย่างเช่น ResNet[1] , VGG[2]

### 2.2. Object Detection

เป็นการตรวจจับวัตถุในรูปภาพด้วยเทคนิคการใช้ Deep Convolutional Neural Network โดยส่วนประกอบภายในของ Object Detection มีการมีใช้ classification model เป็นตัว backbone เพื่อใช้ในการสกัด Feature แต่เพิ่มในส่วนของการทำนายการตีกรอบ หรือ bounding box มาเพื่อให้โมเดลสามารถทำการตรวจจับและหาตำแหน่งของวัตถุ และสามารถทำการตรวจจับได้ ซึ่งโมเดลที่นิยมใช้ในการ Detection ได้แก่ YOLO[8] , Fast RCNN[9]

### 2.3. Transfer learning

การทำ Transfer learning คือ การถ่ายทอดความรู้จากโมเดลหนึ่งมาอีก model หนึ่ง โดยส่วนมากจะเป็นการ Transfer จากโมเดลที่รับการเทรนจาก dataset ใหญ่ๆ โดยการนำ weight จากโมเดลที่ทำการเทรนแล้วนำมาให้อีกโมเดลหนึ่งในการเทรนใหม่เพื่อให้เกิดการประหยัดเวลาในการเทรน และสามารถทำให้โมเดลสามารถเรียนรู้จาก dataset ที่มีขนาดไม่ใหญ่มากได้ โดย Dataset ที่ส่วนมากคนนิยมนำมาใช้ในการ transfer ได้แก่ ImageNet[10] เนื่องจากมีข้อมูลในปริมาณที่มาก.

### 2.4. Fine grain Image Classification

โดยปกติแล้วการทำ Fine grain Image Classification เป็นขั้นเซตของการทำ Image Classification ซึ่งงานของ Fine grain Image Classification นั้นจะเป็นการทำจำแนกวัตถุชนิดต่างๆ ที่มีลักษณะที่คล้ายกันมากๆ เช่น ชนิดของเครื่องบิน ชนิดสายพันธุ์ของสัตว์ เป็นต้น ซึ่งโมเดลในการจำแนกโดยทั่วไปนั้นไม่สามารถแยกออกมาได้อย่างชัดเจน

### 3. Literature Review

#### 3.1. Squeeze and Excitation Networks[5]

โดยเนื้อหาภายในตัวของงานวิจัยได้กล่าวถึง Attention mechanism เทคนิคการเพิ่มตัว SE Block เข้าต่อหลังจากการผ่านในแต่ CNN layer เพื่อการปรับเทียบค่าน้ำหนัก weight ในระหว่าง Channel ซึ่ง SE Block[5] มีการทำงาน 2 ส่วนหลัก ส่วนแรก คือ Squeeze คือการใช้ในการทำการลดมิติลงเหลือ  $1 \times 1 \times \text{Channel}$  มิติ โดยใช้ Global Average Pooling[3] เพื่อเปรียบเทียบค่าความสำคัญของแต่ละชั้น และเข้าต่อในส่วนที่ 2 Excitation เข้าข้อมูลที่ได้มาผ่าน neural network 1 ชั้นเพื่อให้ได้ตัว factor สำหรับการปรับเทียบความสำคัญในแต่ละ Channel และนำตัว factor นี้ไปทำการคูณกับตัวของรูปภาพในแต่ละ channel เพื่อให้เพิ่มความสามารถให้โมเดลสามารถเรียนรู้ส่วน Feature ที่สำคัญ และ ช่วยลดการเรียนรู้ Feature ในส่วนที่ไม่สำคัญลงได้ อีกทั้งตัวของ SE Block ยังสามารถนำไปใช้กับใช้กับตัวโมเดล state of the art ต่างๆได้ เพื่อเพิ่มประสิทธิภาพของโมเดล เช่น SEResNet50

#### 3.2. Bilinear CNNs for Fine-grained Visual Recognition[6]

โดยเนื้อหาภายในตัวของงานวิจัยได้กล่าวถึง การใช้ตัวโมเดล 2 ตัวในการรวม Feature เพื่อใช้ในการจำแนกวัตถุ ซึ่งสามารถทำให้จำแนกวัตถุได้มีประสิทธิภาพมากขึ้น ทำให้สามารถทำการเรียนรู้ตัว Feature ที่มีความละเอียดมากขึ้น ซึ่งในส่วนของการรวม feature ของทั้ง 2 โมเดลนั้น สามารถทำได้โดยการใช้ เทคนิค outer product ในการรวมกันของ feature map ทำให้ได้ bilinear vector และสามารถนำ bilinear vector นั้นในการผ่าน FC layers เพื่อใช้จำแนกวัตถุต่อไปได้ ซึ่งโมเดลนี้สามารถใช้ในการจำแนกวัตถุที่มีลักษณะคล้ายกันหรือใกล้เคียงกันได้

#### 3.3. See Better Before Looking Closer: Weakly Supervised Data Augmentation Network for Fine-Grained Visual Classification หรือ WSDAN[4]

เนื้อหาภายในการกล่าวถึงการใช้เทคนิคการ Augment Data โดยในงานวิจัยได้มีกล่าว ในการ Augment Data แบบปกติโดยเฉพาะ การ Augment แบบ Cropping และ Dropping ซึ่งเทคนิคนี้โดยปกติเป็นการตัดรูปภาพบางส่วนออก หรือ การตัดภาพเฉพาะส่วนนั้นๆออกมาแบบสุ่ม ทำให้เกิดการเรียนรู้ในโมเดลที่ดีขึ้น แต่ทางงานวิจัยได้กล่าวถึงว่า การ Augment แบบนั้นมีโอกาสที่จะทำให้ไม่ได้ช่วยในการเรียนรู้ที่มากนัก เนื่องจากว่า การ Augment นั้นเป็นแบบสุ่มซึ่งอาจจะทำให้การ Augment นั้นไปทำการสุ่มตัดภาพออกมา โดนที่ภาพนั้น ไม่ได้มีรูปของวัตถุอยู่เลย หรือ ทำการสุ่มตัดภาพตัววัตถุออกไปด้วยซ้ำ ทางงานวิจัยจึงได้ใช้ เทคนิคที่ชื่อว่า Attention Cropping และ Attention Dropping เพื่อให้สามารถทำการ Augment ส่วนของรูปภาพได้เหมาะสมมากขึ้น โดยการใช้สิ่งที่เรียกว่า Attention maps มาใช้ในการตรวจหาวัตถุ และ ทำการ Attention Dropping และ Cropping ในส่วนนั้นเพื่อให้โมเดลสามารถเรียนรู้ feature ต่างๆ ของตัววัตถุได้ดียิ่งขึ้น อีกทั้งในการทดสอบก็เช่นกัน ตัวงานวิจัยนี้ได้ใช้ส่วนของ Attention maps มาใช้ใน localization ตัว object เพื่อให้เห็นตัวของวัตถุได้มากยิ่งขึ้น สามารถทำให้ตัวโมเดลนี้สามารถจำแนกวัตถุได้ดี

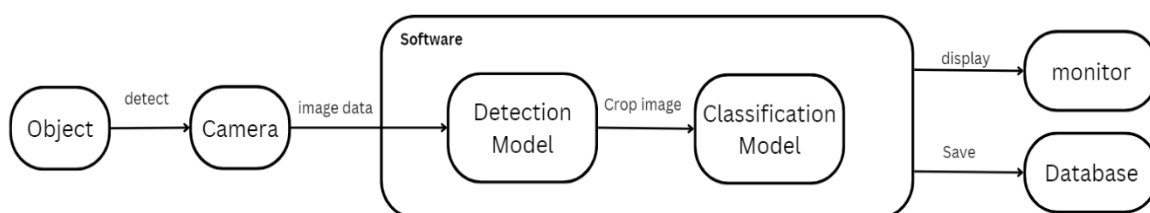
## 4. System design

### 4.1. System requirement

- 4.1.1. ต้องการ accuracy ในการทำนายชนิดแยกชนิดของเครื่องมือผ่าตัดให้ได้ถูกต้องมากกว่า 97 %
- 4.1.2. สามารถเก็บบันทึกข้อมูลหลังการทำนายใน database และสามารถแสดงผลได้
- 4.1.3. สามารถทำงานได้คอยตรวจจับและจำแนกเครื่องมือผ่าตัดได้อย่าง real-time

### 4.2. System overview

โดยระบบมีภาพรวม คือ มีการใช้ตัวของกล้องในการจับภาพวัตถุ และ ทำการส่งข้อมูลรูปภาพนั้นไปให้ในส่วนของ model ต่อๆไปเพื่อใช้ในการทำนายชนิดของอุปกรณ์นั้นๆ



รูปที่ 1 รูปแสดง System overview

- 4.2.1. Detection Model เป็นโมเดลที่ใช้ในการตรวจจับวัตถุเพื่อให้รู้ว่าวัตถุนั้นอยู่บริเวณส่วนไหนของภาพ ทำการตัดภาพเอาภาพของวัตถุทั้งหมดออกมาเป็นรูปภาพย่อยและทำการส่งภาพเหล่านั้นไปให้โมเดลใช้ในการจำแนก เพื่อการรบกวนของพื้นหลังของวัตถุ เพื่อโมเดลสามารถจำแนกชนิดของข้อมูลได้มีประสิทธิภาพมากขึ้น
- 4.2.2. Classification model เป็นโมเดลที่คอยรับข้อมูลรูปภาพย่อยจาก Detection model เพื่อนำมาใช้ในการจำแนกชนิดประเภทของเครื่องมืออุปกรณ์ทางการแพทย์
- 4.2.3. Software เป็นผู้ที่คอยใช้ตัวโมเดลภายในในการจำแนกชนิดของเครื่องมือ และตรวจสอบจำนวนและความถูกต้องของเครื่องมือผ่าตัด และทำการบันทึกข้อมูลลง Database และแสดงผลทาง monitor.

## 5. Resources

### 5.1. Team

Name	Student ID	Task
Tanut Bumrungvongsiri	63340500026	Classification Model
Phufa Boonchuatrong	63340500048	Detection Model

ตารางที่ 1 ตารางแสดงรายชื่อผู้จัดทำ

## 5.2. Task

5.2.1. Classification Model: ทำการพัฒนาโมเดลสำหรับการจำแนกชนิดของเครื่องมือแพทย์

5.2.2. Detection Model: ทำการพัฒนาโมเดลสำหรับการตรวจจับเครื่องมือแพทย์

## 5.3. Process

No	Task	Start Date	End Date
1	เรียนรู้เกี่ยวกับการทำ image classification	3/1/2023	3/7/2023
2	Scrum 1	3/8/2023	
3	ค้นหา Dataset เพื่อนำมาใช้ในการทดลอง Classification	3/9/2023	3/13/2023
4	ทดลองใช้ state of the art model ในการ classification	3/14/2023	3/21/2023
5	Scrum2	3/22/2023	
6	ค้นหา Research Paper ที่เกี่ยวข้อง	3/23/2023	3/30/2023
7	ทดลองใช้ Data augmentation และ Attention mechanism	3/31/2023	4/18/2023
8	Scrum3	4/19/2023	
9	ใช้ Model สำหรับการ Fine Grain Classification	4/20/2023	5/9/2023
10	Scrum 4	5/10/2023	
11	สรุปผลและทำรายงาน	5/11/2023	5/23/2023
12	Final	5/24/2023	

ตารางที่ 2 ตารางแสดง process การทำงาน

## 6. Methodology

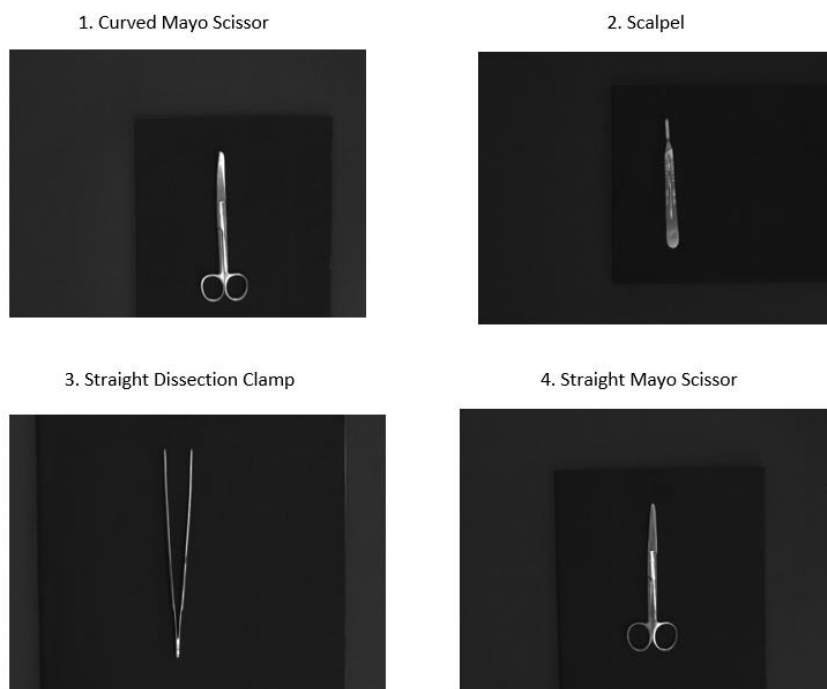
### 6.1. Dataset<sup>[7]</sup>

โดย dataset ที่ผู้จัดทำได้เลือกมาเป็น dataset แบบ open จากตัวเว็บไซต์ Kaggle ซึ่งเป็น dataset ของเครื่องมือผ่าตัด ที่ใช้สำหรับการทำ image classification ซึ่งมีจำนวน class ทั้งหมด 4 class ประกอบด้วยรูปภาพทั้งหมด 2010 รูป

No.	Class	จำนวน
1.	Curved Mayo Scissor	550
2.	Scalpel	550
3.	Straight Dissection Clamp	460
4.	Straight Mayo Scissor	450
รวม		2010

ตารางที่ 3 ตารางแสดง จำนวนข้อมูล

#### 6.1.1. โดยแต่ละคลาสมีลักษณะดังนี้



รูปที่ 2 รูปแสดง ลักษณะของข้อมูลชนิดต่างๆ

- ดยจะสามารถสังเกตเห็นได้ว่าลักษณะของคลาส Curved Mayo Scissor กับ Straight Mayo Scissor นั้นมีความใกล้เคียงกันมาก แตกต่างกันแค่ส่วนปลายของอุปกรณ์
- ผู้จัดทำจึงได้นำโมเดลที่ใช้สำหรับการ Fine grain classification มาใช้ในการทดสอบด้วย

6.1.2. โดยผู้จัดทำได้ทำการแบ่ง dataset ของรูปไว้ใช้ในการ Train : Validate : Test

ในอัตราส่วน 0.7 : 0.1 : 0.2 ซึ่งจะมีจำนวนดังนี้

- Train                      1407    รูป
- Validate                  201    รูป
- Test                        402    รูป

## 6.2. Data Augmentation

6.2.1. Resize ทำการ resize รูปภาพที่ได้รับให้อยู่ใน shape (224,224)

6.2.2. Horizontal flip 0.5 เพื่อเพิ่มมุมมองให้ตัวโมเดลได้เรียนรู้ โดยใช้สำหรับตัว train เท่านั้น

## 6.3. Transfer Learning

6.3.1. โดยโมเดลที่นำมาใช้ในการ train จะทำการ Transfer weight จาก ImageNet[10]

## 6.4. Model

โมเดลที่นำมาใช้ในการทำ Image Classification จะแบ่งการทำทดสอบออกเป็น 2 ประเภท

6.4.1. Model for normal classification

- ResNet50[1]
- VGG16[2]
- SeResNet50 [5]

6.4.2. Model for fine grain classification

- Bilinear ResNet50[6]
- WSDAN[4]

## 6.5. Evaluation

6.5.1. Accuracy : ความถูกต้องของการจำแนกทั้งหมดของโมเดล

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

6.5.2. Sensitivity: ค่าวัดความสามารถของโมเดลที่ใช้ในการทำนายนั้นๆ ได้ถูกขนาดไหน

$$Sensitivity = \frac{TP}{TP + FN}$$

6.5.3. F1-score : ค่าวัดความสามารถของโมเดลในงานการจำแนกที่มีความสำคัญทั้งความแม่นยำ (precision) และความครอบคลุม (recall)

$$F1 \text{ score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

## 7. Results

### 7.1. เปรียบเทียบ Accuracy ในแต่ละโมเดล

Model	Accuracy
ResNet50	0.930
VGG16	0.960
SEResNet50	0.950
Bilinear ResNet50	0.955
<b>WSDAN</b>	<b>0.985</b>

ตารางที่ 4 ตารางแสดง Accuracy ในแต่ละโมเดล

### 7.2. เปรียบเทียบ F1 score ในแต่ละโมเดลและคลาส

Model \ Classes	Curved Mayo Scissor	Scalpel	Straight Dissection Clamp	Straight Mayo Scissor
ResNet50	0.88	0.99	0.99	0.88
VGG16	0.92	0.98	0.98	0.92
SEResNet50	0.91	1.00	1.00	0.90
Bilinear ResNet50	0.93	1.00	0.98	0.92
<b>WSDAN</b>	<b>0.96</b>	<b>1.00</b>	<b>1.00</b>	<b>0.96</b>

ตารางที่ 5 ตารางแสดง F1 score ในแต่ละโมเดลและคลาส

ในแต่ละโมเดลสามารถจำแนกอุปกรณ์ Scalpel และ Straight Dissection Clamp ได้ดีมาก ซึ่งแตกต่างจากอีก 2 ชนิดนั่นก็คือ Curved Mayo Scissor และ Straight Mayo Scissor เพราะเนื่องจากว่าอุปกรณ์ทั้ง 2 ชนิดนี้มีลักษณะที่ใกล้เคียงกันมากทำให้ทำการจำแนกได้ยาก ซึ่งโมเดลที่สามารถทำการจำแนก คือ WSDAN[4] เพราะเป็นโมเดลสำหรับงานด้านนี้โดยเฉพาะ



### 7.3. เปรียบเทียบ Sensitivity ในแต่ละ models และคลาส

<b>Model \ Classes</b>	Curved Mayo Scissor	Scalpel	Straight Dissection Clamp	Straight Mayo Scissor
ResNet50	0.95	0.98	1.00	0.80
VGG16	0.90	0.97	0.99	0.95
SEResNet50	0.93	1.00	1.00	0.88
Bilinear ResNet50	0.95	1.00	0.99	0.89
<b>WSDAN</b>	<b>0.97</b>	<b>1.00</b>	<b>1.00</b>	<b>0.94</b>

ตารางที่ 6 ตารางแสดง Sensitivity ในแต่ละโมเดลและคลาส

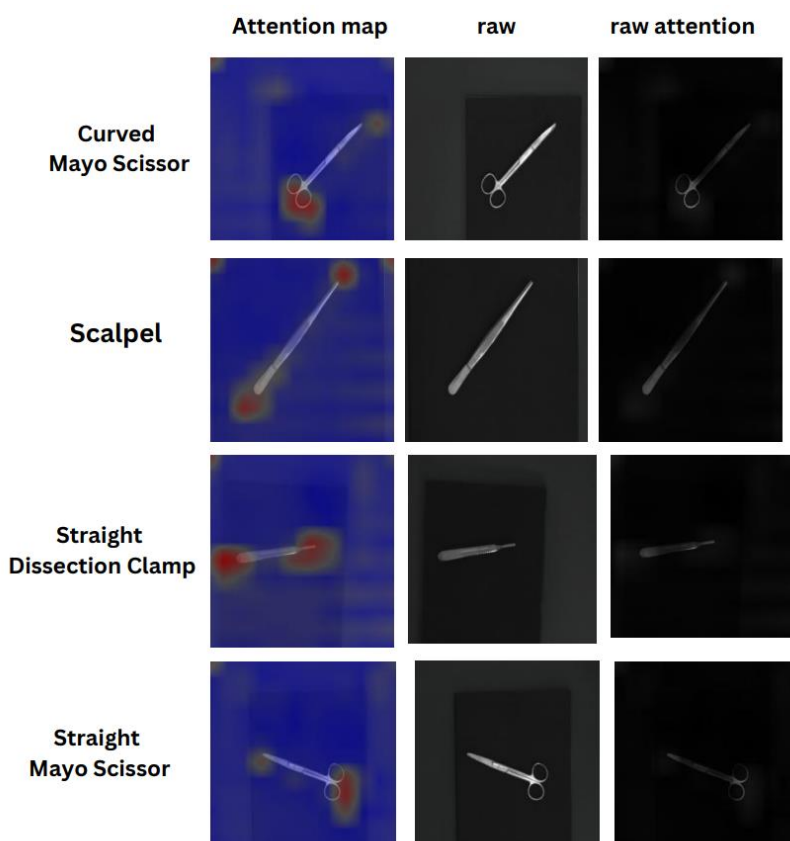
ตารางนี้แสดงให้เห็นถึงความสามารถในการจำแนกของตัวโมเดลว่ามีการจำแนกเครื่องมือประเภทนั้นๆ ถูกมากขนาดไหนและไม่จำแนกผิดเป็นอุปกรณ์ชนิดอื่น ซึ่งจากตารางเห็นได้ว่า โมเดลส่วนมากได้ทำการทำนายอุปกรณ์ Straight Mayo Scissor ถูกต้องน้อยกว่าตัวของอุปกรณ์ชนิดอื่นๆ

## 8. Discussion and Conclusion

### 8.1. Conclusion

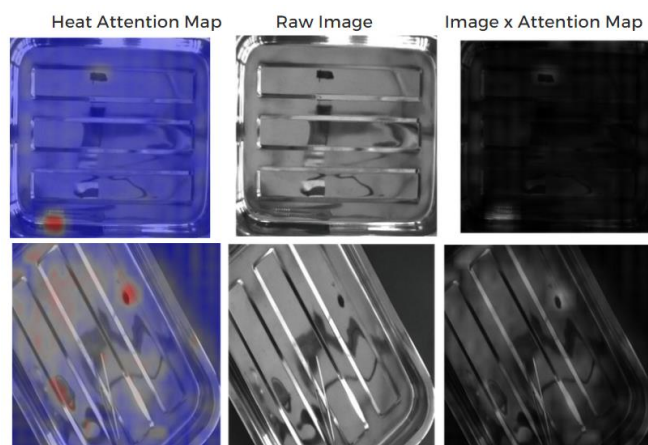
- 8.1.1. โมเดล WSDAN[4] สำหรับการทำ Fine grain classification มีประสิทธิภาพในการจำแนกที่ดีกว่าโมเดลอื่นๆ เพราะเนื่องจากตัวของ Dataset ที่ทำการเลือกมามีชนิดประเภทของกรรไกรผ่าตัดทั้ง 2 แบบ ซึ่งมีลักษณะทางกายภาพที่ใกล้เคียงกันมาก
- 8.1.2. ตัวชนิดของกรรไกรชนิด Straight Mayo Scissor เป็นเครื่องมือที่โมเดลสามารถทำนายได้ถูกน้อยที่สุดถ้าดูจากค่า Sensitivity
- 8.1.3. โมเดลส่วนมากสามารถทำนายเครื่องมือ scalpel กับ Straight Dissection Clamp ได้ดีกว่าเครื่องมืออื่นๆ เพราะอาจจะมีลักษณะที่แตกต่างจากเครื่องมืออื่น
- 8.1.4. โมเดลที่เพิ่ม Attention Mechanism อย่าง SEResNet-50[5] มีประสิทธิภาพในการจำแนกที่ดีมากขึ้นจากตัว ResNet-50[1] ธรรมดา ซึ่งเป็นไปตามงานวิจัยของ SE Block ได้กล่าวไว้
- 8.1.5. โมเดล Bilinear ResNet-50[6] แม้จะเป็นโมเดลที่ทำการรวม โมเดลทั้ง 2 ไว้ด้วยกัน ทำประสิทธิภาพได้ไม่ดีเท่าที่โมเดลอื่นๆ ก็อาจจะเพราะว่าโมเดล Bilinear ResNet-50 อาจจะไม่ได้เหมาะกับตัว Dataset หรือ ไม่ก็ผู้จัดทำอาจจะทำความเข้าใจในตัวโมเดลได้ไม่ดีพอ และยังไม่ได้ลองการรวมโมเดลอื่นๆอีก

## 8.2. Visualization of Attention heat map จากตัวโมเดล WSDAN



รูปที่ 3 รูปแสดง Attention heat map

- 8.2.1. จาก Attention map ที่ได้แสดงให้เห็นว่าตัวของ WSDAN[4] สามารถทำการ localize วัตถุในภาพได้ดีในระดับหนึ่งในสภาพที่ไม่มีพื้นหลัง
- 8.2.2. จากในส่วนของ raw attention เป็นการนำ Attention map รวมกับ raw เปรียบเสมือนภาพที่โมเดลใช้ในการจำแนก
- 8.2.3. ซึ่งก็มีบางภาพที่โมเดลยังไม่สามารถที่จะทำการ attention ตัววัตถุได้ถูกต้องเนื่องจากตัวรูปภาพนั้นมีเงาสะท้อนของตัววัตถุอยู่ทำให้โมเดลเกิดความสับสนได้



รูปที่ 4 รูปแสดง Attention heat map ที่ผิดพลาดของโมเดล

### 8.3. Discussion

- 8.3.1. ในการทำโครงการนี้ ทางผู้จัดทำได้เน้นไปทางการหาโมเดลใหม่ๆ มาเพื่อใช้ในการเรียนรู้ให้มากขึ้น ซึ่งในแต่ละโมเดลนั้นอาจจะสามารถเพิ่มประสิทธิภาพของการจำแนกได้มากกว่านี้อีกถ้ามีการทำ Hyper parameter tuning ซึ่งผู้จัดทำได้ข้ามตรงส่วนนี้ไปเนื่องจากไม่มีเวลามากพอ ถึง โฟกัสไปที่การเรียนรู้โมเดลใหม่ๆ แทน
- 8.3.2. ในโครงการยังไม่ได้มีการศึกษา และ พัฒนาระบบ Data base เลย
- 8.3.3. อีกทั้งยังไม่ได้มีการพัฒนาในเรื่องของการทำให้ Realtime-classification
- 8.3.4. ซึ่งโครงการนี้มีเพียงส่วนเดียวเท่านั้นยังไม่รวมอีกส่วนคือ Object detection

## Reference

- [1]. K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
- [2]. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in ICLR, 2015.
- [3]. M. Lin, Q. Chen, S. Yan, "Network In Network," , National University of Singapore, Singapore, 2013.
- [4]. T. Hu, H. Qi, Q. Huang, Y. Lu, "See Better Before Looking Closer: Weakly Supervised Data Augmentation Network for Fine-Grained Visual Classification", University of Chinese Academy of Sciences, Beijing.
- [5]. J. Hu, L. Shen, S. Albanie, G. Sun and E. Wu, "Squeeze-and-Excitation Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 8, pp. 2011-2023, 1 Aug. 2020, doi: 10.1109/TPAMI.2019.2913372.
- [6]. T. -Y. Lin, A. RoyChowdhury and S. Maji, "Bilinear CNN Models for Fine-Grained Visual Recognition," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 1449-1457, doi: 10.1109/ICCV.2015.170.
- [7]. D. Lavado. (2018). *Labeled Surgical Tools and Images* [Online]. Available: <https://www.kaggle.com/datasets/dilavado/labeled-surgical-tools>
- [8]. J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.
- [9]. R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.
- [10]. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," arXiv:abs/1409.0575 , 2015