

ข้อเสนอโครงการนิเทศน์ ไฟฟ้า วิชา 2102490

การวิเคราะห์ลำดับการสวิงกอล์ฟโดยใช้การเรียนรู้เชิงลึก

Deep Learning Based Golf Swing Sequencing Analysis

นิสิตคนที่ 1

นายคริษฐ์

สุขปรีดี

6230055921

นิสิตคนที่ 2

นางสาวกิงรัก

ไพรاء

6230040421

อาจารย์ที่ปรึกษา

รศ.ดร.สุภาวดี

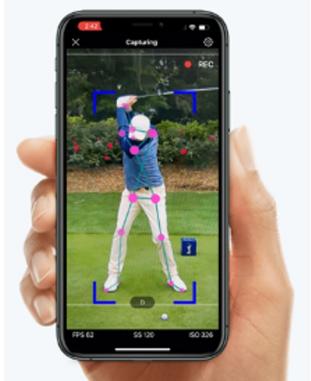
อร่ามวิทย์

ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ปีการศึกษา 2565

หัวข้อนำเสนอ

- ที่มาและความสำคัญของโครงการ
- วัตถุประสงค์ของโครงการ
- ขอบเขตของโครงการ
- หลักการและทฤษฎีที่เกี่ยวข้อง
- ขั้นตอนการดำเนินการ
- ผลลัพธ์จากการดำเนินงานเบื้องต้น
- สรุปภาพรวมของโครงการ
- สิ่งที่คาดว่าจะทำต่อไป
- Q&A

ที่มาและความสำคัญของโครงการ



- ค่าใช้จ่ายที่จำเป็นต่อการเล่นกอล์ฟค่อนข้างสูง เช่น เสื้อผ้า รองเท้า สนามไดร์ฟกอล์ฟ โดยเฉพาะผู้ฝึกสอน
- ปัจจุบันมีโปรแกรมฝึกสอนให้ผู้เล่นได้ฝึกฝนมากมาย ในรูปแบบ IoT หรือ อ�플ิเคชัน แต่ก็ยังคงมี ค่าใช้จ่ายที่สูงอยู่
- ในปัจจุบันยังไม่เคยมีโปรแกรมการวิเคราะห์ท่าสวิงได้ที่ ถูกพัฒนาขึ้นโดยคนไทย และไม่มีค่าใช้จ่ายเลย

ที่มาของรูป:

<https://www.sportsbox.ai>

<https://www.swingtuneup.com>

วัตถุประสงค์ของโครงการ

- นำแบบจำลองการเรียนรู้เชิงลึก (Deep learning model) มาประยุกต์ใช้เข้ากับขั้นตอนวิธีการจับคู่ແຜ່ນແບບໂດຍໃຊ້ມູນ (Template matching algorithm) ທີ່ໄດ້ອອກແບບເອງໃນການວິເຄາະໜໍາດັບການສົງກອລົກ
- ทดสอบແລະປັບປຸງປະສົງປະຕິທີ່ພາບຂອງການໃຊ້ຈຳນວນຂອງໂປຣແກຣມການວິເຄາະໜໍາດັບການສົງກອລົກ



<https://www.youtube.com/watch?v=VopGSOsi-Mc&t=361s>

ขอบเขตของโครงการ

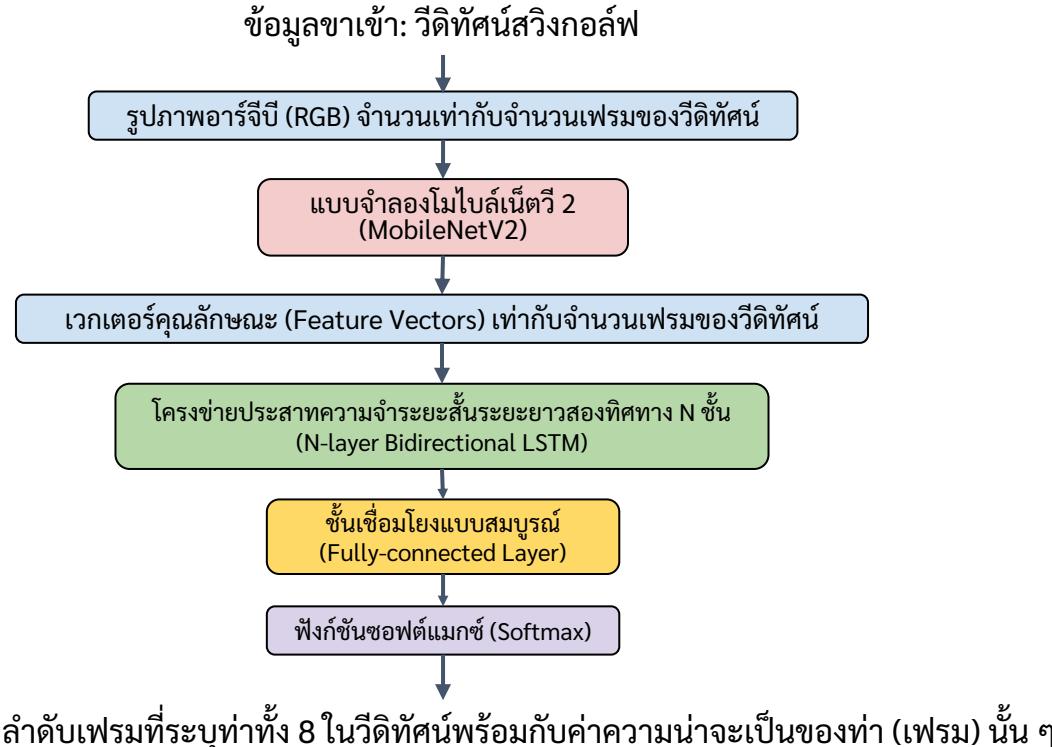
- ใช้ชุดข้อมูลกอล์ฟดีบี (GolfDB) ที่พิจารณาเฉพาะวีดิทัศน์การสวิงกอล์ฟที่แสดงมุมมองด้านหน้า (Front-view) ของผู้เล่นจำนวน 214 วีดิทัศน์
- โครงการนี้จะพัฒนาระบบการวิเคราะห์ลำดับการสวิงกอล์ฟ โดยมีข้อกำหนดดังนี้
 - ใช้แบบจำลองการเรียนรู้เชิงลึกในการจำแนกท่าสวิงทั้ง 8 ท่าในลำดับการสวิงซึ่งจะรับวีดิทัศน์รับเข้าที่มีขนาด 160×160 พิกเซล
 - ใช้แบบจำลองการเรียนรู้เชิงลึกในการตรวจจับจุดสำคัญต่าง ๆ บนร่างกาย 17 จุด ซึ่งจะรับภาพรับเข้าที่มีขนาด 256×256 พิกเซล
 - ใช้ขั้นตอนวิธีการจับคู่แผ่นแบบในการคำนวณคะแนนความคล้ายของแต่ละท่าสวิงในหน่วยเปอร์เซ็นต์
- โปรแกรมนี้จะแสดงผลเป็นคะแนนความคล้ายของแต่ละท่าสวิงในหน่วยเปอร์เซ็นต์

หลักการและทฤษฎีที่เกี่ยวข้อง

1. งานวิจัย: การตรวจจับท่าการสวิงกอร์ฟ
2. งานวิจัย: การตรวจจับจุดสำคัญบนร่างกาย
3. งานวิจัย: การตรวจสอบความคล้ายของการประมาณท่า

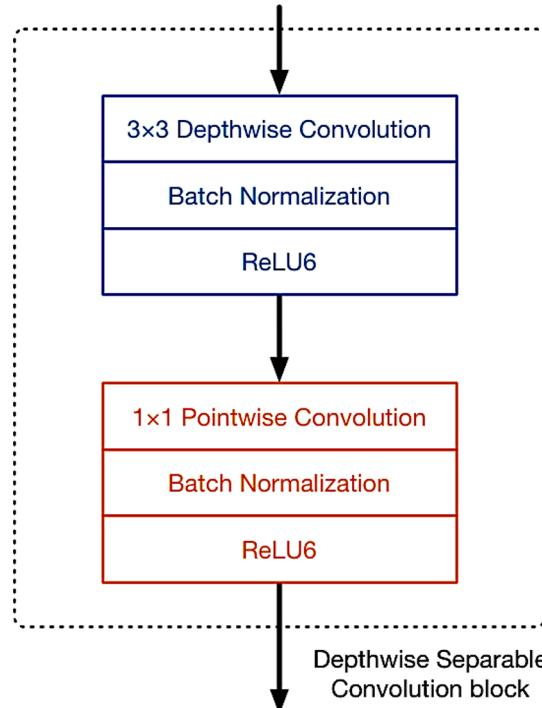
1. งานวิจัยที่เกี่ยวข้องกับการตรวจจับท่าการสวิงกอล์ฟ

แบบจำลองสวิงเน็ต (SwingNet Model)



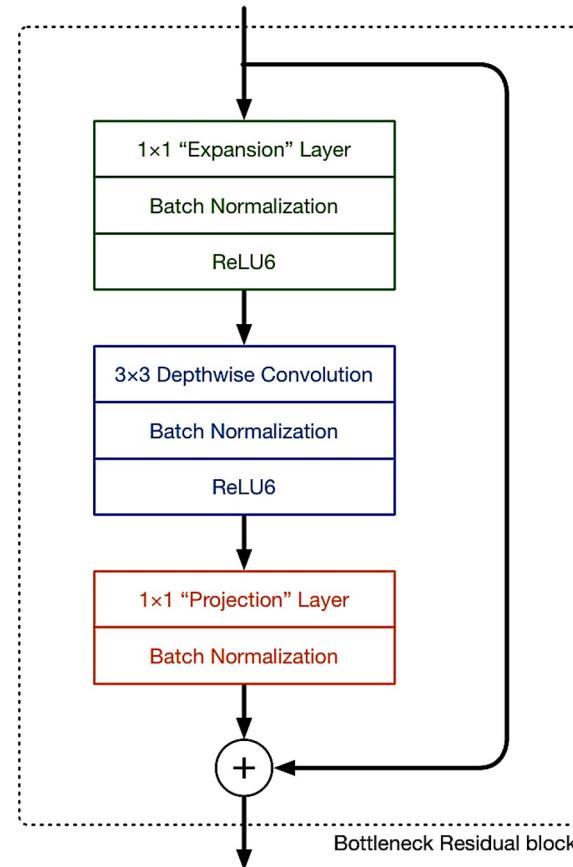
ข้อมูลขาออก: ลำดับเฟรมที่ระบุท่าทั้ง 8 ในวิดีโอทัศน์พร้อมกับค่าความน่าจะเป็นของท่า (เฟรม) นั้น ๆ

โมไบล์เน็ตวี 2 (MobileNetV2)



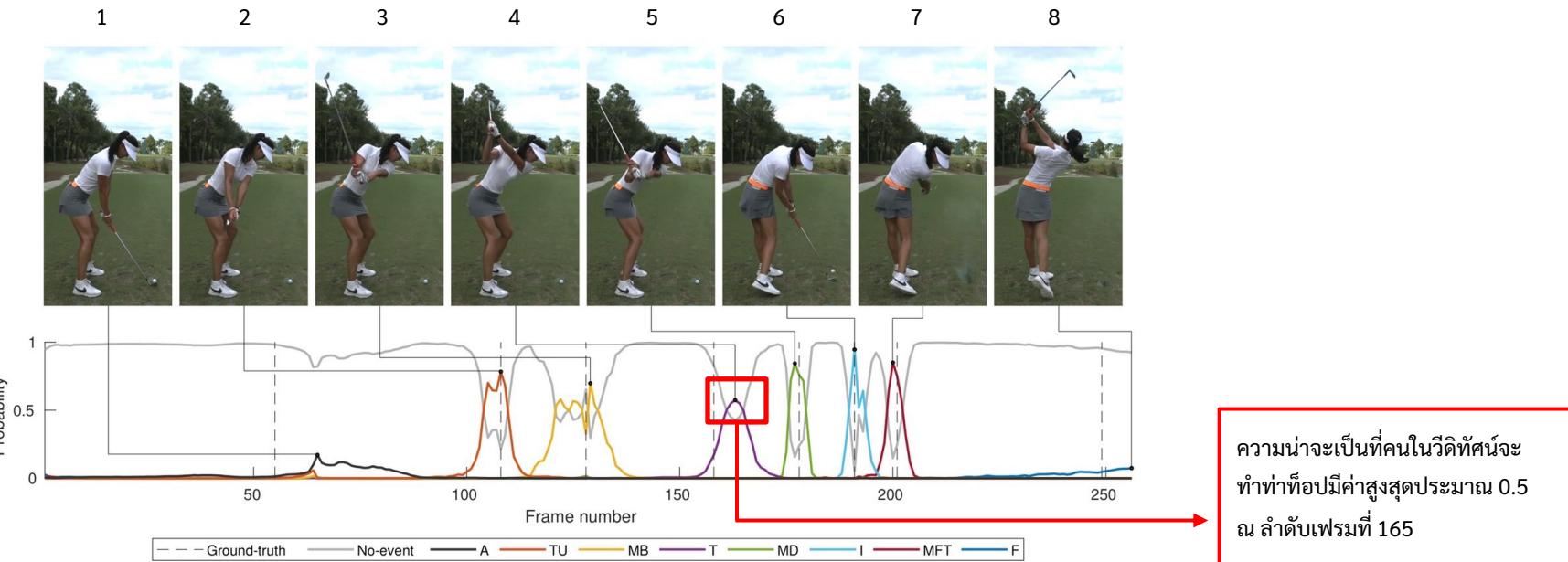
แบบจำลองโมไบล์เน็ตวี 1

เพิ่มการผูกพันของส่วนตกลังและการเปลี่ยนขนาดแบบคอขวดเชิงเส้น (Inverted Residual Bottleneck Layer) & การตัดrelu 6 (ReLU6)



แบบจำลองโมไบล์เน็ตวี 2

ข้อมูลของของแบบจำลองสวิงเน็ต

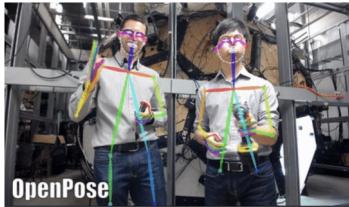


ชื่อท่าสวิงกอล์ฟทั้ง 8 ท่า

- | | | | |
|-------------------------------------|-------------------------|--|----------------------|
| 1. แอ็ดเดรส (Address: A) | 2. โทวอัพ (Toe-up: TU) | 3. มิดแบ็กสวิง (Mid-backswing: MB) | 4. ท็อป (Top: T) |
| 5. มิดดาวน์สวิง (Mid-downswing: MD) | 6. อิมแพกต์ (Impact: I) | 7. มิดฟอลโลว์ทรู (Mid-follow-through: MFT) | 8. พินิช (Finish: F) |

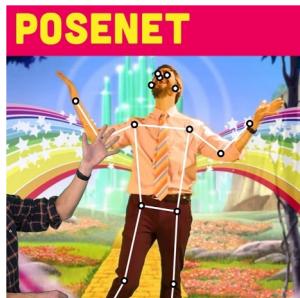
2. งานวิจัยเกี่ยวกับการตรวจจับจุดสำคัญบนร่างกาย (Keypoints Detection)

โอเพนโพส (OpenPose)



<https://paperswithcode.com/paper/openpose-realtime-multi-person-2d-pose>

โพสเน็ต (PoseNet)



<https://www.youtube.com/watch?v=Olo-DlOkNVg>

มูฟเน็ต (MoveNet)



<https://github.com/CMU-Perceptual-Computing-Lab/openpose>

OpenPose: Cao, Zhe, et al. "Realtime multi-person 2d pose estimation using part affinity fields." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

PoseNet: G. Papandreou, T. Zhu, L.-C. Chen, S. Gidaris, J. Tompson, and K. Murphy. Personlab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model. In Proceedings of the European Conference on Computer Vision (ECCV), pages 269–286, 2018.

MoveNet: Jo, BeomJun, and SeongKi Kim. "Comparative Analysis of OpenPose, PoseNet, and MoveNet Models for Pose Estimation in Mobile Devices." Traitement du Signal 39.1 (2022): 119-124.

เวลาที่ใช้ (วินาที) ในการประมาณท่าของรูปทั้งหมด
1,000 รูปของแต่ละกลุ่ม

Group	OpenPose		PoseNet	
	Total time	SD	Total time	SD
1st	649.368	13.757	78.289	0.252
2nd	643.384	21.915	77.732	0.630
3rd	637.856	16.960	69.675	0.464

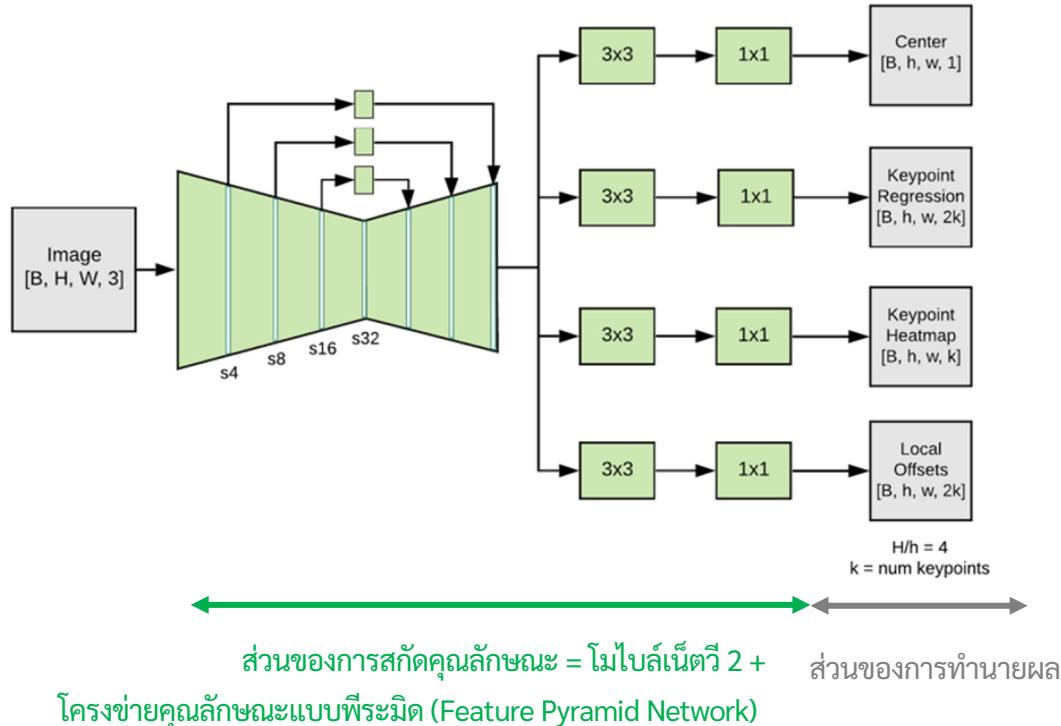
Group	MoveNet Lightning		MoveNet Thunder	
	Total time	SD	Total time	SD
1st	55.362	0.652	141.435	8.186
2nd	56.122	1.086	141.005	3.832
3rd	48.891	1.746	137.482	0.717

ความถูกต้องของการประมาณท่าของรูปในกลุ่มที่ 1 และ 3

Model	OpenPose	PoseNet	MoveNet Lightning	MoveNet Thunder
Group 1	78.5%	96.7%	78.7%	79.5%
Group 3	93.8%	98.4%	71.5%	81.6%
Average	86.2%	97.6%	75.1%	80.6%

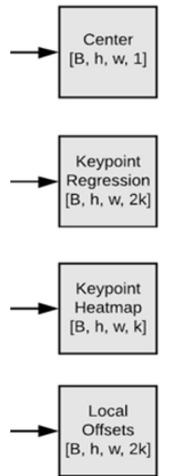
- จากการวิจัยในปี 2022 สามารถสรุปได้ว่า
 - ความถูกต้อง: **โพสเน็ต > โอลเคนโพส > มูฟเน็ต**
 - ความเร็วในการประมาณท่า: **มูฟเน็ต (ไลต์นิง) > โพสเน็ต > โอลเ肯โพส**
- แต่เลือกใช้แบบจำลองมูฟเน็ต เพราะใช้งานง่าย และปรับแต่ง/ต่อ�อดโปรแกรมเองได้สะดวกกว่า

โครงสร้างของแบบจำลองมูฟเน็ต



- ผลลัพธ์จะแสดงจุดสำคัญ 17 จุด + ค่าความเชื่อมั่น (กำหนดโดยชุดข้อมูลโคโค (COCO Keypoint Dataset Training & Validation Set 2017))

ส่วนการทำนายผลของแบบจำลองมูฟเน็ต



แผนภูมิความร้อนของ
จุดสำคัญ (Keypoint
Heatmap)



ค่าดัดเชยของจุดสำคัญ
(Local Offsets)

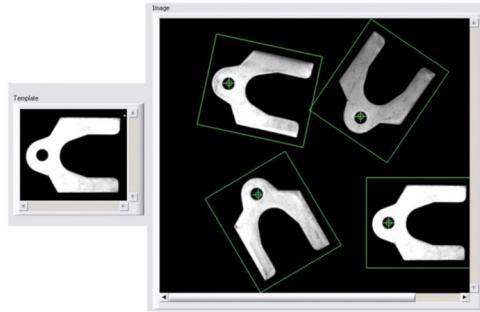


แบบจำลองมูฟเน็ตทั้งสองชุด

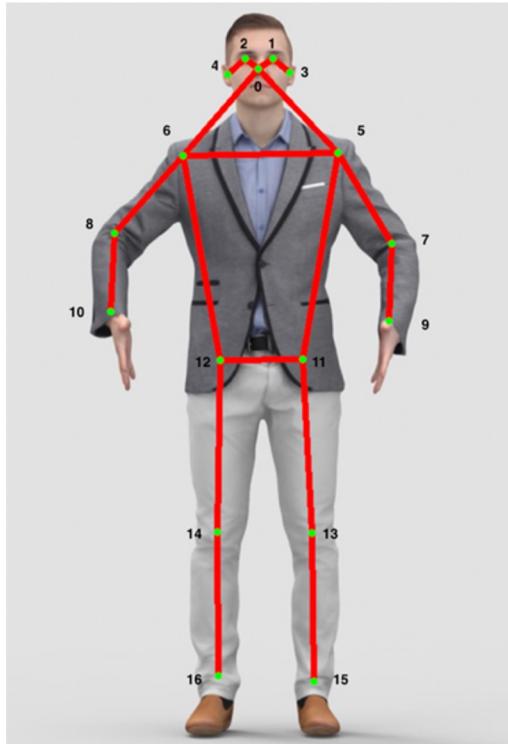
ชุด (Version)	ขนาดของวีดิทศน์หรือภาพรับเข้า [พิกเซล]	ตัวคูณความลึก (Depth Multiplier)	เวลาที่ใช้ทดสอบ (Inference Time)	ค่า Mean Average Precision
ไลต์นิง (Lightning)	192 x 192	1.0	ต่ำ	ต่ำ
ทันเดอර์ (Thunder)	256 x 256	1.75	สูง	สูง

- ★ แบบจำลองมูฟเน็ตถูกฝึกฝนและทดสอบโดยชุดข้อมูลโคโค (COCO Keypoint Dataset Training & Validation Set 2017) และชุดข้อมูลภายนอก (Active Dataset Training & Evaluation Set)

3. งานวิจัยเกี่ยวกับการตรวจสอบความคล้ายของการประมาณท่า (Pose Estimation)



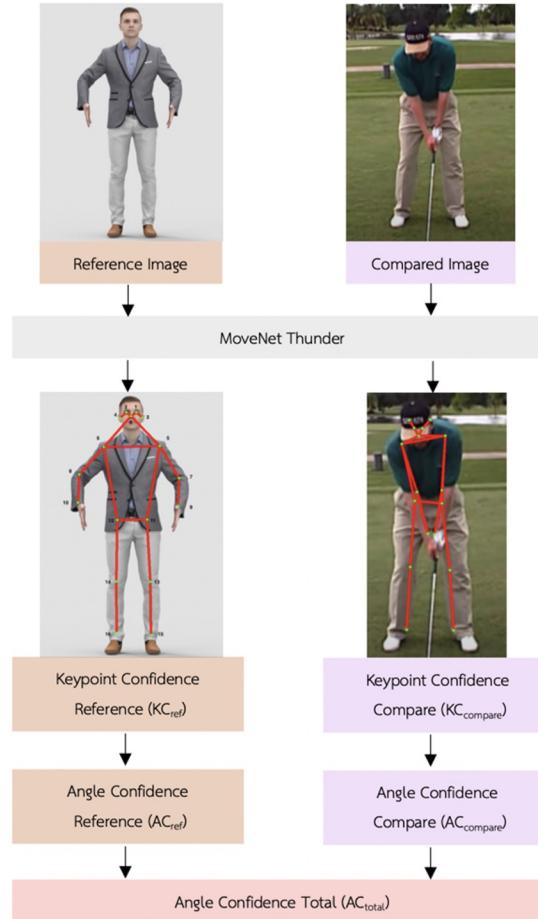
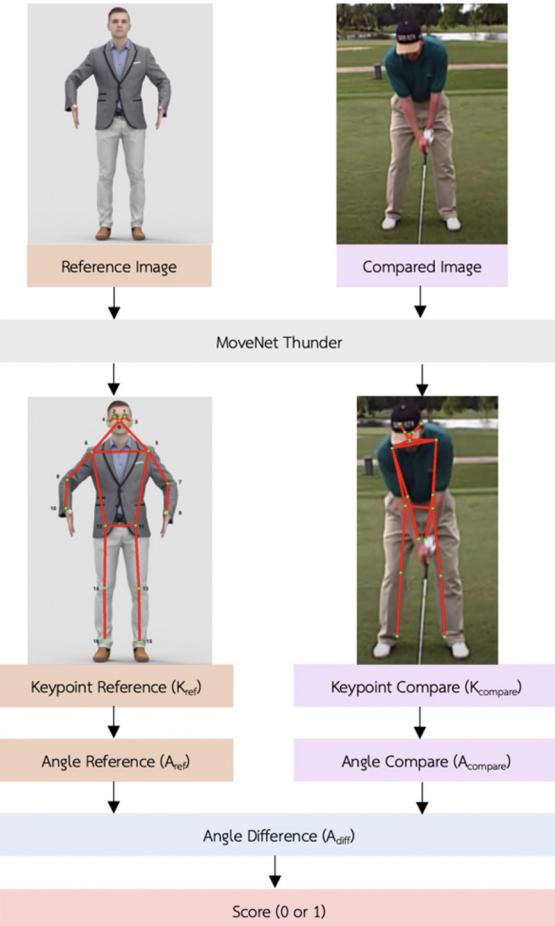
General Template Matching Algorithm



จุดสำคัญทั้งหมด 17 จุดหลังผ่านแบบจำลองมูฟเน็ตทันเดอร์

- ขั้นตอนวิธีการจับคู่แผ่นแบบ (Template Matching Algorithm)
- ปัญหา: วัตถุที่จะนำมาค้นหาต้องมีลักษณะเหมือนกัน แต่การสวิงกล้องนั้นไม่สามารถเทียบกันได้ เพราะเป็นคนละคน และสวมคนละชุด
- การแก้ปัญหา: ใช้แนวคิดเรื่องมุมของแต่ละจุดสำคัญเป็นตัวเปรียบเทียบกันแทน โดยได้แรงบันดาลใจจากงานวิจัย Match Pose ในการออกแบบโปรแกรมต่อไป

แนวคิดของขั้นตอนวิธีการจับคู่แผ่นแบบโดยใช้มุมที่ออกแบบเอง



การคำนวณคะแนนความคล้าย
(Score of Similarity: SOS)

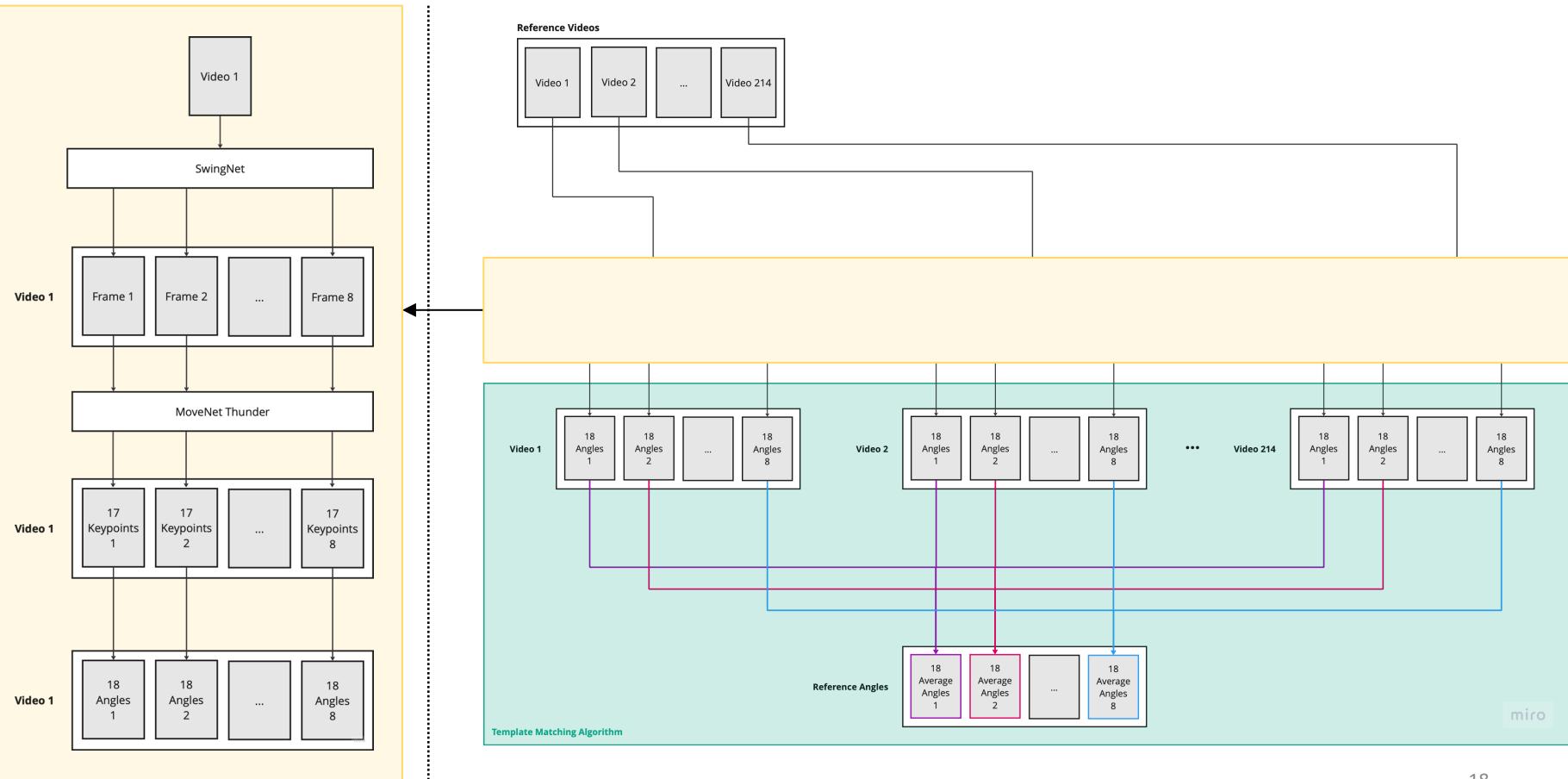
$$SOS = \frac{\sum_{n=0}^{17} (c_n s_n)}{\sum_{n=0}^{17} (c_n)} \times 100\%$$

เมื่อ n คือ ลำดับของมุม
 c_n คือ ค่าความเชื่อมั่นรวมของมุมที่ n
 s_n คือ ค่าคะแนน (0 หรือ 1) ของมุมที่ n

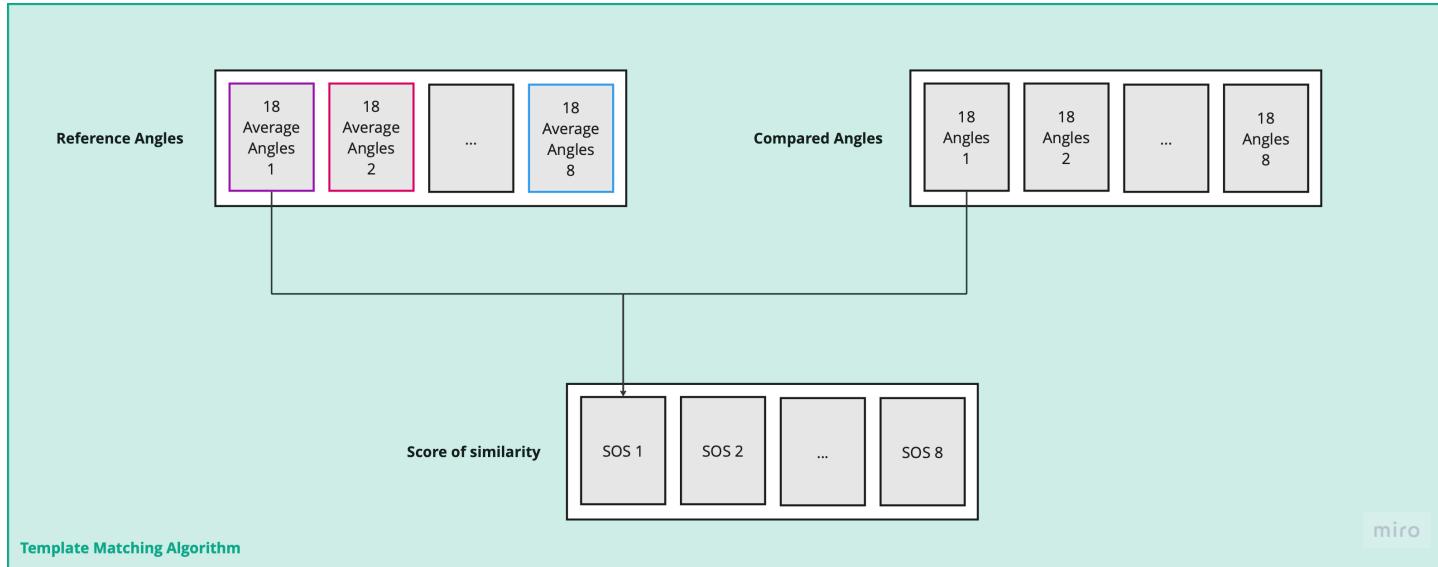
ผลลัพธ์การดำเนินงานเบื้องต้น

1. ภาพรวมของขั้นตอนการดำเนินงาน
2. การจัดการชุดข้อมูลกอล์ฟดีบี
3. ทดสอบการทำงานและผลลัพธ์ที่ได้จากแบบจำลองสวิงเน็ต
4. ทดสอบการทำงานและผลลัพธ์ที่ได้จากแบบจำลองมูฟเน็ตทันเดอร์
5. ทดสอบการทำงานและผลลัพธ์ที่ได้จากขั้นตอนวิธีการจับคู่แผ่นแบบโดยใช้มุม

1. ภาพรวมของขั้นตอนการดำเนินงาน: การเตรียม Reference Angles



1. ภาพรวมของขั้นตอนการดำเนินงาน: การคำนวณคะแนนความคล้าย (SOS)



2. การจัดการชุดข้อมูลกอล์ฟดีบี (GolfDB)

id	youtube_id	player	sex	club	view	slow	events	bbox	split
9	#NAME?	CRISTIE KERI	f	driver	face-on	1	[601 631 69C [3.90625000		
11	gOBVh7fzY2l	STEVE STRICK	m	driver	face-on	1	[744 803 8 [1.06250000		
15	vN3Uc_Ehnn	GREG NORRIM	m	driver	face-on	1	[809 833 8 [6.01562500		
20	iW323nsTGT	HYO JOO KIM	f	driver	face-on	1	[578 592 65C [0.084375 C		
22	xD6KDqPF9c	TIGER WOODS	m	iron	face-on	1	[738 755 80C [0.07890625		
26	pxO_eGmIDt	PORNAVONI	f	driver	face-on	1	[742 776 8 [0.06953125		
28	2Us22X_S6h	ANGEL YIN	f	iron	face-on	1	[483 555 58C [0.1671875 I		
30	B1uIW4LN1e	BRANFORD	m	driver	face-on	1	[807 875 9 [0.0984375 I		
44	j6h-3Ez0714	COLIN MONTE	m	driver	face-on	1	[2168 2204 2 [7.42187500		
52	kSCTyo10cnL	BROKE HEN	f	driver	face-on	1	[1012 1080 1 [5.15625000		
61	TH7DFdUgQvKA	NA JANG	f	driver	face-on	1	[765 811 84C [6.09375000		
67	8n9BBNEj3yI	LYDIA KO	f	iron	face-on	1	[1281 1283 1 [1.49218750		
75	UZ9f9suLkG3	INBEE PARK	f	driver	face-on	1	[1048 1127 1 [0.07890625		
85	Wee5Vhf-iv	ANGEL YIN	f	iron	face-on	1	[576 601 63C [1.62500000		
99	NhB8KnoufLi	PORNAVONI	f	iron	face-on	1	[927 986 1C [0.1546875 I		
103	r9Nfuhb0aYv	IN-GEE CHI	f	driver	face-on	1	[653 669 74C [0.08359375		



- วิดิทัศน์การสวิงกอล์ฟโดยนักกีฬากอล์ฟมืออาชีพทั้งหมด 1,400 วิดิทัศน์ที่ถูกตัด (trim) ให้ทุกวิดิทัศน์แสดงเฉพาะช่วงการสวิงเท่านั้นและยังมีการตัด (crop) ให้แสดงแต่ผู้เล่นเป็นหลัก
- แต่ละแบบจำลองที่ใช้ในโครงงานมีความต้องการขนาดของวิดิทัศน์รับเข้า (input video) ไม่เท่ากัน เราจึงได้ปรับแต่งโปรแกรมจากงานวิจัย และเขียนโปรแกรมเพิ่มเติมเพื่อที่จะปรับแต่งขนาดวิดิทัศน์ได้
- เราคัดเลือกเฉพาะวิดิทัศน์ที่แสดงมุมมองด้านหน้าของผู้เล่นทั้งสิ้น 214 วิดิทัศน์มาเท่านั้น

3. ทดสอบการทำงานและผลลัพธ์ที่ได้จากแบบจำลองสวิงเน็ต

- 1.) ผลทดสอบประสิทธิภาพของแบบจำลองที่ผ่านการเรียนรู้แล้ว (Pretrained Model) มีความสอดคล้องกับงานวิจัยจริง มีการกำหนดพารามิเตอร์ Input Size = 160, Sequence length = 64, Batch = 6 ,LSTM layers = 2, LSTM Hidden = 128, Params = 4.07×10^6 , FLOPs = 10.75, Learning Rate = 0.01
- 2.) ผลการทดสอบด้วยวิดีทัศน์ทดสอบของงานวิจัย โดยใช้วิดีทัศน์สวิงกอล์ฟของ มิเชล ไว (Michelle Wie) ในการแบ่งขั้นแอลพีจีเอ

		ความน่าจะเป็น (สูงสุด) ที่คนในวิดีทัศน์กำลังทำท่าหนึ่ง ๆ อยู่							
		แออดเดรส	ໄทเรอัพ	มิดแบ็กสวิง	ท้อป	มิดดาวน์สวิง	อิมแพกต์	มิดฟอลโลว์thrū	พินิช
ลำดับบุฟเฟรมจากการทดสอบ เองที่ปรากฏค่าความน่าจะ เป็นนั้น ๆ		0	63	86	116	128	139	146	206
ที่มาของ ผลลัพธ์	งานวิจัย	0.200	0.750	0.600	0.500	0.750	0.850	0.800	0.100
	ทดสอบ	0.089	0.662	0.480	0.091	0.919	0.921	0.908	0.045
ผลต่าง		0.111	0.088	0.120	0.409	0.169	0.071	0.108	0.055



3.1) ผลการทดสอบด้วยวิดีทัศน์ที่ไม่ได้มาจากงานวิจัย (ในมุมด้านข้าง)

โดยใช้วิดีทัศน์สวิงกอล์ฟของนักกีฬากอล์ฟมืออาชีพ ไทเกอร์ วูดส์ (Tiger Woods)

ความน่าจะเป็น (สูงสุด) ที่คนในวิดีทัศน์กำลังทำท่าหนึ่ง ๆ อยู่

	แอ็อดเดรส	ໂຫວັນ	ມິດແບກສົງ	ທົ່ອປ	ມິດດາວນໍສົງ	ອິມແພເກຕ໌	ມິດຝອລໍລ້ວທຽງ	ພິນີ້
ลำดับเฟรมที่ปราฏถูกค่าความน่าจะเป็นน า	60	39	113	193	209	239	249	306
ผลการทดสอบ	0.057	0.534	0.552	0.242	0.624	0.866	0.728	0.084



ภาพที่ได้จากการทดสอบมุมด้านข้างเรียงลำดับตามตารางข้างต้น

3.2)

ผลการทดสอบด้วยวีดิทัศน์ที่ไม่ได้มาจากงานวิจัย (ในมุมด้านหน้า)

โดยใช้วีดิทัศน์สวิงกอล์ฟของนักกีฬากอล์ฟมืออาชีพ ไทเกอร์ วูดส์ (Tiger Woods)

ความน่าจะเป็น (สูงสุด) ที่คนในวีดิทัศน์กำลังทำท่าหนึ่ง ๆ อยู่

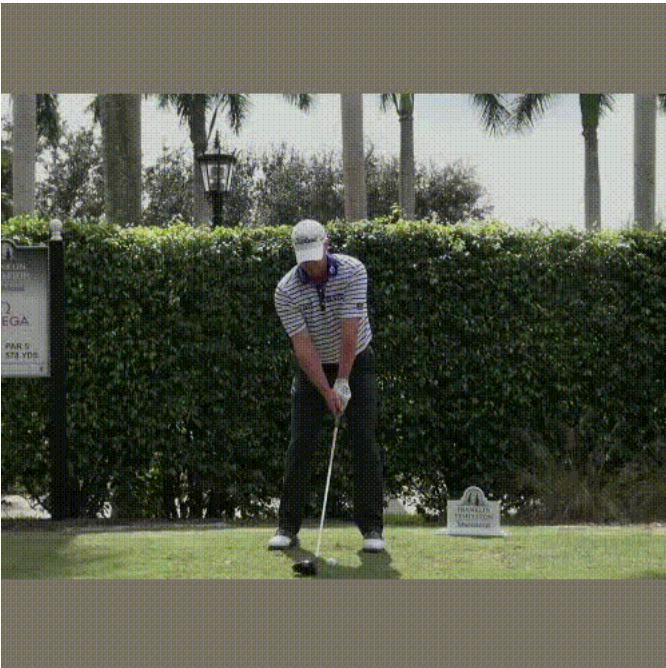
	แออดเดรส	โหวอัพ	มิดแบ็กสวิง	ท็อป	มิดดาวน์สวิง	อิมแพกต์	มิดฟอลโลว์ทรู	พินิช
ผลการทดสอบจากมุมด้านหน้า	0.011	0.753	0.956	0.617	0.807	0.977	0.854	0.314
ผลการทดสอบจากมุมด้านข้าง	0.057	0.534	0.552	0.242	0.624	0.866	0.728	0.084
ผลด้านหน้า - ผลด้านข้าง	- 0.046	0.219	0.404	0.375	0.183	0.111	0.126	0.230



ภาพที่ได้จากการทดสอบวีดีโอมุมด้านหน้าเรียงลำดับท่าตามตารางข้างต้น

4. ทดสอบการทำงานและผลลัพธ์ที่ได้จากแบบจำลองมูฟเน็ตทันเดอร์

วีดิทัศน์ A



<https://www.youtube.com/watch?v=gOBVh7fvz7o>

วีดิทัศน์ B



https://www.youtube.com/watch?v=vN3Uc_EhnnY

4. ทดสอบการทำงานและผลลัพธ์ที่ได้จากแบบจำลองมูฟเน็ตทันเดอร์

วีดิทัศน์ A					วีดิทัศน์ B													
เฟรมที่	รูป	จุดสำคัญที่ได้จากแบบจำลองมูฟเน็ตทันเดอร์	จุดสำคัญที่นำไปพิสูจน์ในโปรแกรมแม่พากล็อตลิบ	รูป	จุดสำคัญที่ได้จากแบบจำลองมูฟเน็ตทันเดอร์	จุดสำคัญที่นำไปพิสูจน์ในโปรแกรมแม่พากล็อตลิบ	รูป	จุดสำคัญที่ได้จากแบบจำลองมูฟเน็ตทันเดอร์	จุดสำคัญที่นำไปพิสูจน์ในโปรแกรมแม่พากล็อตลิบ									
1		0.207			0.207				0.150									
2		0.446			0.446				0.150									
3		0.651			0.651				0.194									
4		0.513			0.513				0.908									
5		0.879			0.879				0.348									

5. ทดสอบการทำงานและผลลัพธ์ที่ได้จากขั้นตอนวิธีการจับคู่แผ่นแบบโดยใช้มุ่ง

- ★ จะใช้ผลลัพธ์ของเฟรมทั้ง 5 คู่จากวิดีทัศน์หน้าที่แล้วมาใช้ทดสอบ

ตารางที่ 1: มุมทั้ง 18 มุมของทั้ง 5 เฟรมของวิดีทัศน์ A

เฟรมที่/มุมที่	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	30.02	41.88	110.56	128.81	82.22	80.49	1.68	6.44	177.40	165.96	96.45	100.84	172.80	151.87	90.75	107.29	179.25	176.19
2	29.29	46.09	79.70	161.19	92.32	75.87	41.91	39.24	155.17	155.09	84.96	106.86	174.74	150.80	100.30	102.34	176.61	171.35
3	50.00	40.87	77.47	160.27	76.99	94.29	49.51	64.58	128.29	135.00	93.62	95.11	179.16	155.45	87.23	109.44	169.17	170.60
4	9.69	98.13	152.82	142.13	88.49	91.24	54.64	28.51	8.13	25.35	73.07	107.20	161.22	146.61	88.15	106.19	156.35	163.81
5	16.26	153.43	76.33	87.71	120.41	61.56	60.34	57.30	111.80	120.90	69.89	108.14	170.00	144.98	100.11	106.89	171.97	166.29

ตารางที่ 2: มุมทั้ง 18 มุมของทั้ง 5 เฟรมของวิดีทัศน์ B

เฟรมที่/มุมที่	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	20.35	15.68	94.02	103.17	74.60	86.06	0.93	1.43	166.15	176.07	98.60	100.74	167.12	151.14	94.28	108.12	173.48	175.75
2	5.19	6.12	57.91	119.13	106.33	71.05	53.62	41.96	158.96	157.68	83.09	99.52	175.74	161.95	101.17	98.53	176.40	169.51
3	10.22	37.50	62.97	169.51	95.32	83.17	168.51	69.81	76.43	163.93	84.41	97.10	173.91	163.85	101.68	99.05	173.12	166.87
4	4.40	45.00	26.57	98.13	92.94	87.06	123.90	129.81	172.30	147.43	87.06	92.94	179.36	161.64	93.58	105.42	164.55	159.05
5	8.17	85.91	25.34	105.26	91.92	90.82	74.76	78.02	152.41	169.94	76.87	100.40	178.16	158.91	104.98	100.69	175.84	164.03

ตารางที่ 3: ผลต่างมุมทั้ง 18 มุมของทั้ง 5 เฟรมของวิดีทัศน์ A และ B

เฟรมที่/มุมที่	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	9.67	26.20	16.53	25.64	7.62	5.57	0.76	5.01	11.25	10.11	2.15	0.10	5.68	0.73	3.53	0.83	5.77	0.45
2	24.10	39.97	21.79	42.07	14.02	4.82	11.71	2.73	3.79	2.59	1.87	7.33	1.00	11.15	0.86	3.81	0.21	1.84
3	39.77	3.36	14.50	9.24	18.33	11.11	119.00	5.23	51.86	28.93	9.21	1.99	5.25	8.40	14.45	10.39	3.96	3.73
4	5.29	53.13	126.25	43.99	4.44	4.17	69.26	101.30	164.17	122.08	13.99	14.26	18.13	15.03	5.42	0.77	8.20	4.76
5	8.09	67.52	50.99	17.55	28.49	29.26	14.42	20.72	40.61	49.04	6.97	7.74	8.16	13.94	4.87	6.19	3.87	2.26

5. ทดสอบการทำงานและผลลัพธ์ที่ได้จากขั้นตอนวิธีการจับคู่แฟ้มแบบโดยใช้มุ่ง

ตารางที่ 4: คะແນນທັງ 18 ມຸນຂອງທັງ 5 ເຟຣມຂອງວິດີທັສນີ A ແລະ B ເມື່ອ Threshold = 15°

ເຟຣມທີ/ມຸນທີ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	ຄະແນນຮຽນ
1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	15	
2	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	14	
3	0	1	1	1	0	1	0	1	0	0	1	1	1	1	1	1	1	13	
4	1	0	0	0	1	1	0	0	0	0	1	1	0	0	1	1	1	9	
5	1	0	0	0	0	0	1	0	0	0	1	1	1	1	1	1	1	10	

ตารางที่ 5: คະແນນຄວາມຄ້າຍຂອງທັງ 5 ເຟຣມຂອງວິດີທັສນີ A ແລະ B ເມື່ອ Threshold = 15°

ເຟຣມທີ	ຄະແນນຄວາມຄ້າຍ [%]
1	84.76
2	79.01
3	76.01
4	57.64
5	60.94

- ค่า Threshold ສາມາດປັບຕາມທີ່ຜູ້ໃຊ້ຕ້ອງການກຳຫົດ ດ້ວຍໆຢັ້ງມາກ = ຍິ່ງຕ້ອງພາຍາມເລີຍນໍາໃຫ້ເໜີມືອນກັບ Reference
- ค่า Threshold ມີໄວ້ເພື່ອຮະບຸນຸມທີ່ມີຄວາມແຕກຕ່າງຮ່ວ່າງຜູ້ໃຊ້ແລະ Reference ມາກເກີນໄປ ເພື່ອໃຫ້ຜູ້ໃຊ້ປັບປຸງທ່ານ໌ ຖ.

สรุปภาพรวมของโครงงาน

- ศึกษาพื้นฐานของการเรียนรู้เชิงลึก (Deep Learning)
- ศึกษางานวิจัยที่เกี่ยวข้อง
- ศึกษาและปรับแต่งโปรแกรมในงานวิจัย
- จัดการกับชุดข้อมูล
- วางแผนการดำเนินงานเบื้องต้น
- ออกแบบขั้นตอนวิธีขั้นตอนวิธีการจับคู่แผ่นแบบโดยใช้มุม
- ทดสอบการทำงาน

สิ่งที่คาดว่าจะทำต่อไป

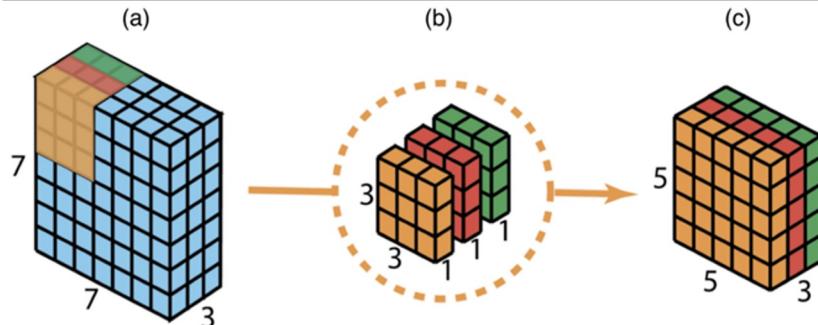
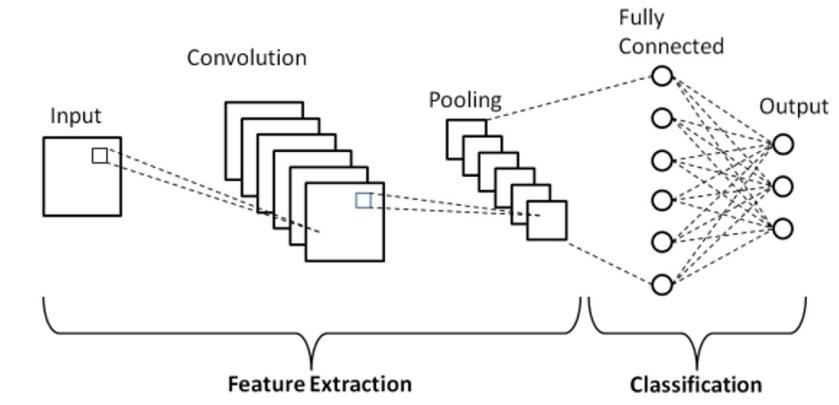
- ศึกษาขั้นตอนวิธีอื่น ๆ ที่จะเพิ่มประสิทธิภาพให้แบบจำลอง
- รวมโปรแกรมเพื่อทดสอบการใช้งานจริง
- ประเมินผลและปรับปรุงแก้ไข

THANK YOU

BACKUP SLIDES

โครงข่ายประสาทแบบสัมภัตนาการ (Convolutional Neural Network, CNN)

พิจารณาส่วนย่อยของภาพนั้นหลาย ๆ ส่วน
ร่วมกันแล้ววิเคราะห์ออกมาเป็นภาพรวมคล้าย
การมองเห็นของมนุษย์ การพิจารณาส่วนย่อยของ
ภาพจะใช้ตัวกรอง (Filter) สัมภัตนาการกับพื้นที่
ย่อยนั้น ๆ และได้ผลลัพธ์ออกมาเป็นผัง
คุณลักษณะ (Feature Map)



- พารามิเตอร์ที่สำคัญ
 - การเสริมเต็ม (Padding)
 - สไทร์ด (Stride)
- นอกจากนี้การสัมภัตนาการยังทำงาน
ประกอบกับกระบวนการอื่นๆ เช่น การ
ดึงคุณลักษณะ (Pooling)

การทำให้เป็นบรรทัดฐานแบบกลุ่ม (Batch Normalization)

ช่วยเพิ่มความเร็วในการฝึกสอนโมเดลและทำให้โมเดลเป็นอิสระต่อกันมากขึ้นในแต่ละชั้นโดยมีสูตรเบื้องต้นดังสมการด้านล่างนี้

$$x^{**} = \left(\frac{x - E[x]}{\sqrt{var(x)}} \right) \gamma + \beta$$

x^{**} คือ ผลลัพธ์สุดท้ายจากการทำให้เป็นบรรทัดฐานแบบกลุ่ม

x คือ ข้อมูลขาเข้า

$E[x]$ คือ ค่าเฉลี่ยของกลุ่มข้อมูลขาเข้า

$var(x)$ คือ ค่าความแปรปรวนของกลุ่มข้อมูลขาเข้า

γ, β คือ ค่าถ่วงน้ำหนักที่ได้จากการเรียนรู้ของชั้นโครงข่ายประสาทนั้น ๆ

ชั้นเชื่อมโยงแบบสมบูรณ์ (Fully-connected Layer)

คือโครงข่ายประสาทที่เชื่อมต่อทุกจุดเชื่อมต่อ (Node) ที่เป็นไปได้กับทุกชั้นโดยจะเชื่อมผังคุณลักษณะที่ได้ออกมาในแต่ละชั้นอย่างสมบูรณ์ด้วยการแปลงแบบเชิงเส้นหรือการคูณเชิงสเกลาร์ (Dot Product) ดังสมการ

$$y_{jk}(x) = f \left(\sum_{i=1}^{n_H} w_{jk} x_i + w_{j0} \right)$$

w_{jk} คือ ค่าถ่วงน้ำหนักในเมตริกซ์ถ่วงน้ำหนักแควที่ j หลักที่ k

x_i คือ เวกเตอร์ของข้อมูลตัวที่ i ที่รับเข้ามา

w_{j0} คือ ค่าถ่วงน้ำหนักความเอนเอียง

relu 6 (Rectified Linear Unit Activation Layer: ReLU6)

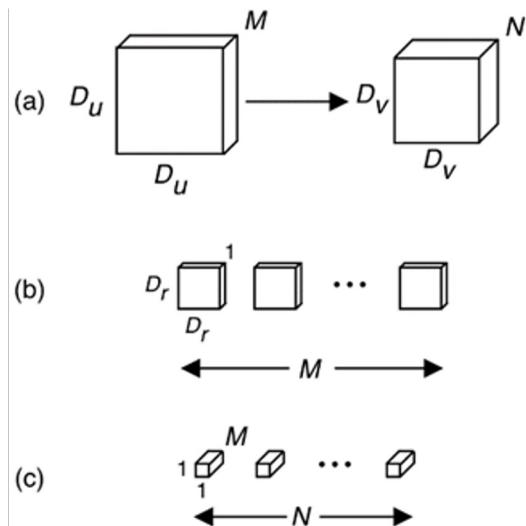
relu คือ พังก์ชันที่ไม่เป็นเชิงเส้นชนิดหนึ่งที่นิยมใช้ในฝึกสอนแบบจำลอง ที่ช่วยให้แบบจำลองสามารถเรียนรู้ได้รวดเร็วและยังเรียนรู้ความซับซ้อนของข้อมูลได้มากขึ้น โดยมีการวิเคราะห์ดังสมการ

$$f(x) = \max(0, x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}$$

โดย relu 6 คือ relu ที่ถูกจำกัดให้การกระตุน (Activation) มีค่าไม่เกิน 6 เพื่อเพิ่มประสิทธิภาพของการคำนวณที่มีความแม่นยำต่อ

การสังวัตนาการชนิดเดพธิไวซ์ (Depth-wise Convolution)

ช่วยลดต้นทุนในการประมวลผลเนื่องจากมีการใช้ตัวกรองสองครั้งทำให้ขนาดของข้อมูลลดลงหลายเท่าเมื่อเทียบกับการสังวัตนาการแบบปกติ



$$D_s = D_r^2 D_v^2 MN$$

$$D_d = D_r^2 D_v^2 M + D_v^2 MN$$

D_s คือ มิติของผลลัพธ์ของการสังวัตนาการชนิดมาตรฐาน

D_d คือ มิติของผลลัพธ์ของการสังวัตนาการชนิดเดพธิไวซ์

D_r คือ มิติของตัวกรองตัวที่หนึ่ง

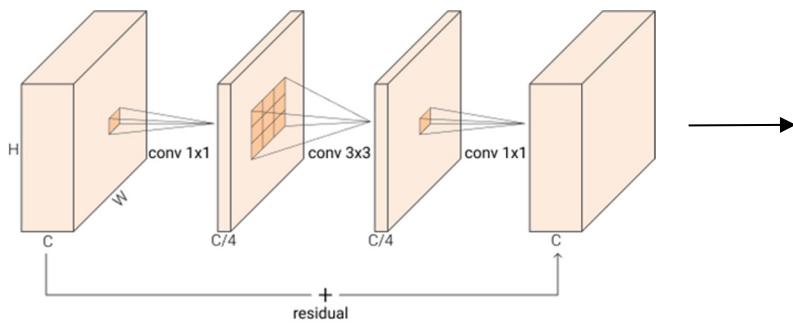
D_v คือ มิติของข้อมูลขาออก

M คือ จำนวนของตัวกรองตัวที่หนึ่ง

N คือ จำนวนของตัวกรองตัวที่สอง

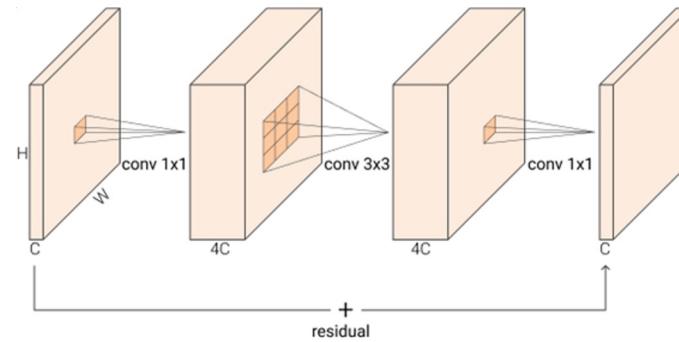
การผกผันของส่วนตอกค้างและการเปลี่ยนขนาดแบบคอขวดเชิงเส้น (Inverted Residual Bottleneck Layer)

การเปลี่ยนขนาดแบบคอขวดเชิงเส้น



กว้าง -> แคบ -> กว้าง

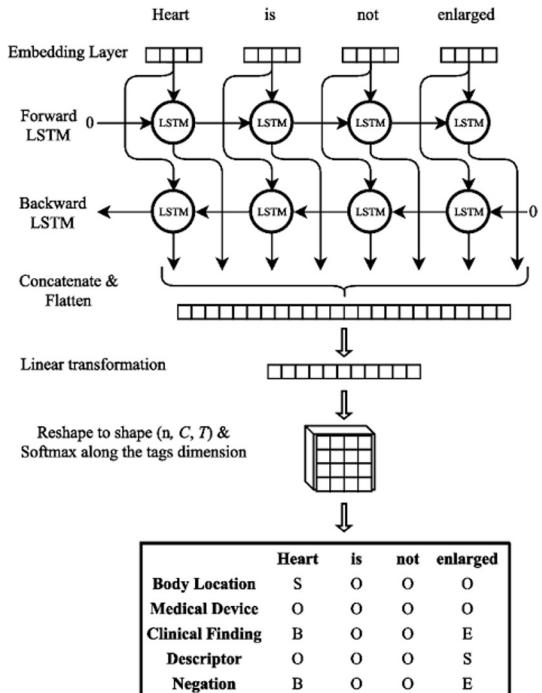
กระบวนการผกผันของส่วนตอกค้างและการเปลี่ยนขนาดแบบคอขวดเชิงเส้น



แคบ -> กว้าง -> แคบ

***ลดปริมาณข้อมูลที่พิจารณาในชั้นต่อไป

โครงข่ายประสาทความจำระยะสั้นระยะยาวสองทิศทาง N ชั้น (N-layer bidirectional LSTM)



โครงข่ายประสาทที่จะถ่วงน้ำหนัก
จากข้อมูลทั้งอดีตและอนาคตซึ่งจะช่วย
เพิ่มประสิทธิภาพในการเรียนรู้ทำให้ถูกใจ
ในการวิเคราะห์ประโยชน์คอมพิวเตอร์ คำพูด หรือการ
ทำนายอนาคตอย่างแพร่หลาย

ฟังก์ชันซอฟต์แมกซ์ (Softmax)

ฟังก์ชันซอฟต์แมกซ์คือฟังก์ชันที่ช่วยปรับค่าของผลลัพธ์ให้เหมาะสมโดยมีค่าอยู่ในช่วง 0 ถึง 1 ซึ่งช่วยให้ง่ายต่อการตีความและยังช่วยลดการมีอิทธิพลของค่าที่สูงหรือต่ำเกินไป ซึ่งมีหลักการคำนวณดังสมการ

$$\sigma(j) = \frac{\exp(w_j^T x)}{\sum_{k=1}^K \exp(w_k^T x)} = \frac{\exp(z_j)}{\sum_{k=1}^K \exp(z_k)}$$

K คือ จำนวนข้อมูลข้าอกอกทั้งหมด

x คือ ข้อมูลข้าอกอกของแต่ละชั้น

w คือ ค่าถ่วงน้ำหนัก

$\sigma(j)$ คือ ผลลัพธ์ที่ได้จากการใช้ฟังก์ชันซอฟต์แมกซ์

ร้อยละความถูกต้องของเหตุการณ์ (Percent Correct Event, PCE)

ร้อยละความถูกต้องของเหตุการณ์คือค่าที่งานวิจัยใช้วัดประสิทธิภาพของแบบจำลองสwingเน็ตเวิร์กการวิเคราะห์ค่าเบื้องต้นเริ่มด้วยการกำหนดค่าที่ยอมรับของผลต่างระหว่างหมายเลขของเฟรมที่ได้จากการทำนายและหมายเลขของเฟรมที่ได้จากฐานข้อมูลจากด้วยสมการ

$$\delta = \max\left(\left\lfloor \frac{n}{f} \right\rfloor, 1\right)$$

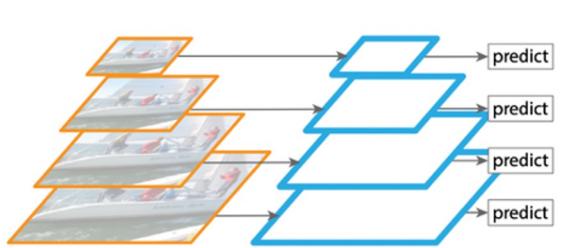
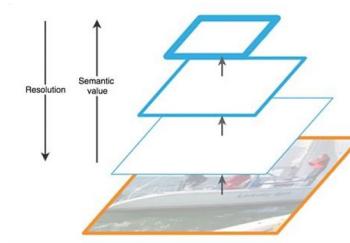
δ คือ ค่าที่ยอมรับของผลต่างระหว่างหมายเลขของเฟรมที่ได้จากการทำนายและหมายเลขของเฟรมที่ได้จากฐานข้อมูล

n คือ จำนวนเฟรมของวีดีโอจากฐานข้อมูลตั้งแต่เอดเดรสสิ่งอิมแพกต์

f คือ อัตราเฟรม โดยมีค่าโดยปริยายเท่ากับ 30 เฟรมต่อวินาที

การสกัดคุณลักษณะ

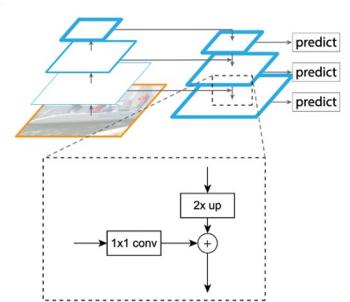
- ภาพที่ผ่านการสังวัตนาการจะมีความละเอียดที่น้อยลงแต่จะมีความหมายมากขึ้น



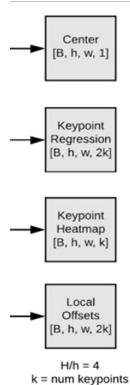
- สิ่งของขนาดเล็ก ๆ จะเป็นที่จะได้ตรวจจับในภาพที่มีความละเอียดสูง แต่สำหรับสิ่งของขนาดใหญ่ ๆ สามารถตรวจจับได้ ในภาพที่มีความละเอียดต่ำ
- เพื่อความแม่นยำจึงควรทำนายในทุก ๆ ชั้นของการสังวัตนาการ?
- แต่จะสิ้นเปลืองเวลาและหน่วยความจำ

โครงข่ายคุณลักษณะแบบพีระมิด (Feature Pyramid Network)

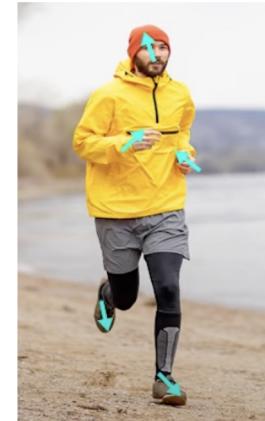
- การสร้างคืนชั้นต่าง ๆ กลับมาในฝั่งของท็อป-ดาวน์ (Top-down) โดยมีการเพิ่มส่วนต่อข้าม (Skip Connection)
- เพื่อลดผลกระทบจากการสูตรตัวอย่างลง (Downsampling) และการสูตรตัวอย่างขึ้น (Upsampling)
- เพิ่มความแม่นยำของตำแหน่งของวัตถุที่ตรวจจับได้มากขึ้น



การทำนายผล



จุดศูนย์กลางของวัตถุ
(Center)
จุดสำคัญที่ถูกคาดถอย
(Keypoint Regression)



แผนภูมิความร้อนของ
จุดสำคัญ (Keypoint
Heatmap)

ค่าชดเชยของจุดสำคัญ
(Local Offsets)

- จุดสำคัญที่ถูกคาดถอยที่ขึ้นกับจุดศูนย์กลางของวัตถุหนึ่ง ๆ จะถูกทำนายออกมาก่อน
- หลังจากนั้นจะทำนายแผนภูมิความร้อนของจุดสำคัญออกมา
- แต่ละพิกเซลของแต่ละแผนภูมิความร้อนของจุดสำคัญจะถูกคูณด้วยค่าถ่วงน้ำหนักที่ขึ้นกับส่วนกลับของระยะห่างระหว่างพิกเซลนั้น และจุดสำคัญที่ถูกคาดถอยที่เกี่ยวข้อง เพื่อจำแนกว่าจุดสำคัญหนึ่ง ๆ เป็นของวัตถุใดในภาพ (ในกรณีที่มีหลายวัตถุในภาพเดียวกัน) หรือก็คือจะเป็นการรับประทานว่าจะไม่สนใจวัตถุอื่น ๆ ที่อยู่ในพื้นหลังที่ไม่ใช่วัตถุหลัก
- อย่างไรก็ตามตำแหน่งของจุดสำคัญจากแผนภูมิความร้อนของจุดสำคัญที่เพิ่มการชดเชยเข้าไปให้ถูกต้องมากขึ้น

การประเมินผลแบบจำลองมุฟเน็ต

- mAP (Keypoint Mean Average Precision) ร่วมกับ OKS (Object Keypoint Similarity)

$$OKS = \exp\left(-\frac{d_i^2}{2s^2k_i^2}\right)$$

เมื่อ d_i คือ ระยะห่างระหว่างจุดสำคัญที่แท้จริงกับจุดสำคัญที่ได้จากการทำนาย
 s คือ ค่าคงที่สำหรับการปรับสเกล k_i คือ ค่าคงที่ของแต่ละจุดสำคัญสำหรับการควบคุมการฟอลloff (Falloff) ในฟังก์ชันเลขชี้กำลัง

- ค่า OKS จะแสดงว่าจุดสำคัญที่ได้จากการทำนายนั้นอยู่ใกล้กับจุดสำคัญที่แท้จริงเท่าไร (ค่า OKS มีค่าตั้งแต่ 0 ถึง 1)
- เราสามารถบอกได้ว่าการตรวจจับจุดสำคัญจะถูกต้องก็ต่อเมื่อจุดสำคัญดังกล่าวมีค่า OKS มากกว่าค่าขีดแบ่ง (Threshold) ค่าหนึ่ง
- ค่า AP จะคำนวนจากการทำนายที่ถูกและผิดของแต่ละจุดสำคัญ (เมตริกซ์สับสน) โดยมีค่า OKS เป็นขีดแบ่งแทนที่จะใช้ค่า IoU (Intersection over Union) แบบวิธีทั่วไป
- สุดท้ายนำมาหาค่าเฉลี่ยของจุดสำคัญทั้งหมดโดยเป็นค่า mAP

		Predicted Class		Sensitivity $\frac{TP}{(TP + FN)}$	Specificity $\frac{TN}{(TN + FP)}$	Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$
		Positive	Negative			
Actual Class	Positive	True Positive (TP)	False Negative (FN) Type II Error			
	Negative	False Positive (FP) Type I Error	True Negative (TN)			
	Precision	$\frac{TP}{(TP + FP)}$	Negative Predictive Value	$\frac{TN}{(TN + FN)}$		