

# Estimación de Pose y Clasificación de Imágenes

<sup>1</sup>Centro de Investigación en Matemáticas A.C.



Centro de  
Investigación  
en Matemáticas, A.C.

# Tabla de Contenido

- 1 Motivación de la Tesis
- 2 Descripción de los Problemas
- 3 Modelos
- 4 Variación a Transformers: Cabezas de Atención Flexibles
- 5 Resultados

# Table of Contents

- 1 Motivación de la Tesis
- 2 Descripción de los Problemas
- 3 Modelos
- 4 Variación a Transformers: Cabezas de Atención Flexibles
- 5 Resultados

# Motivación de la Tesis

Con el auge de los Transformers como modelos de procesamiento de información secuencial, el trabajo de esta tesis ha sido dirigido en explorar dichos modelos en áreas fuera del Procesamiento del Lenguaje Natural.

Finalmente se propone una variante enfocado en aumentar la capacidad receptiva de las cabezas de atención permitiendo mayor flexibilidad al no estar ligada al tamaño de embedding predefinidos.

Las experimentaciones del funcionamiento la variante del modelo se realizan en los siguientes problemas:

- Predicción de Pose 2D en humanos sobre imágenes
- Predicción de Pose 3D en humanos (Monocular, Desacoplado)
- ViT y Clasificación de Enfermedades Comunes de Tórax (INAOE, CIMAT, IMSS)

# Table of Contents

- 1 Motivación de la Tesis
- 2 Descripción de los Problemas
- 3 Modelos
- 4 Variación a Transformers: Cabezas de Atención Flexibles
- 5 Resultados

# Estimación de Pose 2D y 3D en Humanos

- 2D: Dada una imagen estimar las posiciones de las articulaciones de la persona en cuestión sobre la imagen.
- 3D: Dada una imagen estimar las posiciones de las articulaciones dentro de un marco de referencia que mejor ajuste la posición espacial de la persona en cuestión.



Figure 1: Estimación de Pose 2D

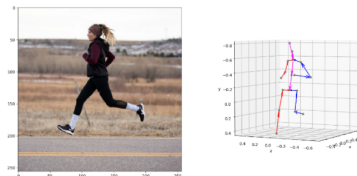


Figure 2: Estimación de Pose 3D

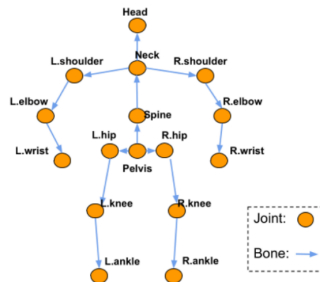
# Estimación de Pose 2D y 3D en Humanos

## Esqueleto

- 17 articulaciones
- 16 huesos

## Datasets

- Human 3.6M
- COCO 2017



**Figure 3:** Articulaciones usadas en tareas de Estimación de Pose en Humanos. Imagen tomada del artículo "Anatomy-Aware 3D Human Pose Estimación en Videos"

# Detección y Clasificación de Enfermedades Comunes de Tórax

- Trabajo colaborativo entre CIMAT, INAOE e IMSS.
- Modelo clasificador para la detección de 15 padecimientos incluyendo COVID-19.
- Se realiza la comparativa de un modelo basado en ViT con las modificaciones antes mencionadas.

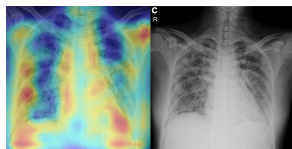


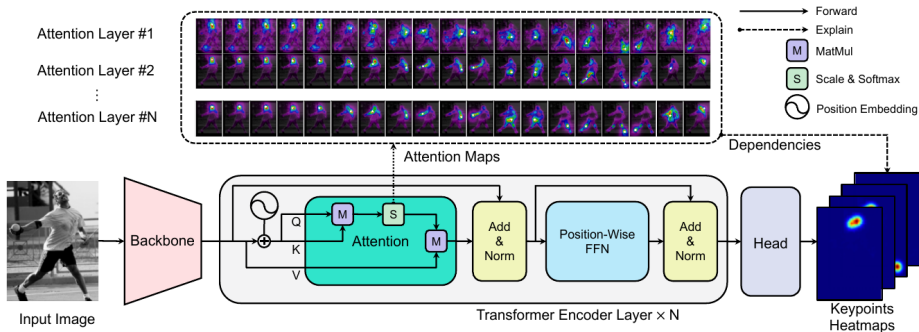
Figure 4: Áreas Afectadas por COVID-2019 detectadas por el modelo usando GradCam.



# Table of Contents

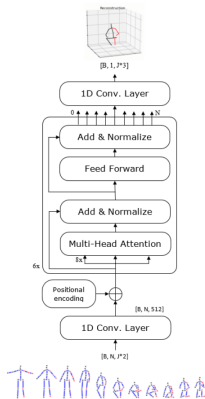
- 1 Motivación de la Tesis
- 2 Descripción de los Problemas
- 3 Modelos**
- 4 Variación a Transformers: Cabezas de Atención Flexibles
- 5 Resultados

# Modelo Estimación de Pose 2D

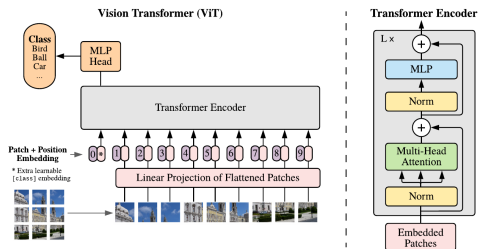


**Figure 5:** Modelo Predicción de Pose 2D. Al igual que ViT usa capas con Decoders con entrada las características obtenidas por un modelo convolucional usado como Backbone

# Modelo Estimación de Pose 3D y ViT



**Figure 6:** Modelo Estimación de Pose 3D. Solo usa capas con Decoders.



**Figure 7:** Modelo ViT usado en las tareas de clasificación de enfermedades pulmonares. La entrada es una secuencia obtenida al dividir la imagen en pequeños parches.

# Table of Contents

- 1 Motivación de la Tesis
- 2 Descripción de los Problemas
- 3 Modelos
- 4 Variación a Transformers: Cabezas de Atención Flexibles
- 5 Resultados



# MultiHead-Self-Attention

El Transformer está basado en la idea de Multihead-Self-Attention (MHA), permitiendo al modelo conjuntamente atender a información en diferentes posiciones desde diferentes subespacios de representación.

$$mha(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \text{head}_3, \dots, \text{head}_h) W^O \quad (1)$$

donde  $Q, K, V \in \mathbb{R}^{n \times d_m}$  son embeddings de entrada,  $n$  es el tamaño de la secuencia,  $d_m$  es el tamaño del embedding y  $h$  el número de cabezas de atención. Cada cabeza es definida como sigue:

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) = \text{Sofmax}\left(\frac{QW_i^Q(KW_i^K)^T}{\sqrt{d_k}}\right) VW_i^V \quad (2)$$

donde  $W_i^Q, W_i^K \in \mathbb{R}^{d_m \times d_k}$ ,  $W_i^V \in \mathbb{R}^{d_m \times d_v}$ ,  $W^O \in \mathbb{R}^{hd_v \times d_m}$  y  $d_k = d_v = d_m/h$

# MultiHead-Self-Attention - Flexible

- El tamaño de la cabeza es dependiente de la dimensión del embedding y el número de cabezas de atención.
- Mientras más cabezas de atención los embeddings son proyectados a dimensiones cada vez más bajas, lo que implica una mayor compresión y pérdida de información.
- Escalar el modelo implica escalar conjuntamente todas las dimensiones.

Redefiniendo  $W_i^Q, W_i^K, W_i^V \in \mathbb{R}^{d_m \times d_h}$ ,  $W^O \in \mathbb{R}^{hd_h \times d_m}$  con  $d_h$  como un hiperparámetro y  $d_h > d_k$  podemos solventar los puntos anteriores.

# Table of Contents

- 1 Motivación de la Tesis
- 2 Descripción de los Problemas
- 3 Modelos
- 4 Variación a Transformers: Cabezas de Atención Flexibles
- 5 Resultados**

Modelo	$d_m$	$d_h$	AP	AR
Base	256	32	0.726	0.780
Propio	256	16	0.737	0.766
Propio	256	32	0.741	0.771
Propio	256	64	0.748	0.776

**Table 1:** Average Precision y Average Recall sobre el conjunto de validación COCO-2017 para Estimación de Pose 2D.

Modelo	$d_m$	$d_h$	Prot. 1	Prot. 2
Base	512	64	37.7mm	27.6mm
Propio	512	32	46.2mm	33.7mm
Propio	512	64	45.15mm	32.96mm
Propio	512	128	45.11mm	32.84mm

**Table 2:** Error en milímetros usando Protocolos 1 y 2 para Estimación de Pose 3D.



# Detección y clasificación de Enfermedades Comunes de Tórax

	Compared Models				Proposal	
	CheXNeXt	CRAL	DR-CNN	TSNC	Original	Relabeled
Cardiomegaly	0.885	0.880	0.801	0.887	0.869	0.920
Emphysema	0.906	0.908	0.773	0.930	0.931	0.959
Effusion	0.825	0.829	0.797	0.831	0.834	0.903
Hernia	0.901	0.917	0.748	0.921	0.856	0.961
Infiltration	0.694	0.702	0.751	0.703	0.712	0.817
Mass	0.824	0.834	0.760	0.833	0.818	0.874
Nodule	0.759	0.773	0.741	0.798	0.795	0.880
Atelectasis	0.769	0.781	0.766	0.785	0.773	0.859
Pneumothorax	0.715	0.729	0.778	0.731	0.895	0.949
Pleural-Thick.	0.766	0.778	0.759	0.782	0.803	0.913
Pneumonia	0.852	0.857	0.800	0.881	0.856	0.925
Fibrosis	0.821	0.830	0.765	0.833	0.854	0.943
Edema	0.842	0.850	0.820	0.849	0.866	0.949
Consolidation	0.745	0.754	0.787	0.754	0.752	0.886
COVID-19	—	—	—	—	0.999	0.999
Healthy	—	—	—	—	0.731	0.873
Average-14	0.807	0.816	0.775	0.823	0.834	0.915
Average	—	—	—	—	0.841	0.916

Table 3: Comparativa de ROC AUC para distintos modelos sobre el conjunto de validación.

# Cronograma de Actividades Restantes

Julio				Agosto				Septiembre			
W1	W2	W2	W4	W1	W2	W2	W4	W1	W2	W2	W4
Últimos Resultados y experimentos											
Redacción de Tesis											
				Revisión y Correcciones							
							Exámen de Grado				

Figure 8: Cronograma de Actividades

Gracias por su atención.



Centro de  
Investigación  
en Matemáticas, A.C.