

EFECTO DE LA TRANSFERENCIA DE
APRENDIZAJE EN LA RED VGG16 PARA EL
RECONOCIMIENTO DEL ALFABETO DE LA
LENGUA DE SEÑAS AMERICANO.

Instituto Tecnológico y de Estudios Superiores de Monterrey.

Participantes:

Regina Alexia Blas Flores.

Emiliano Vivas Rodríguez.

Armando Arredondo Valle.

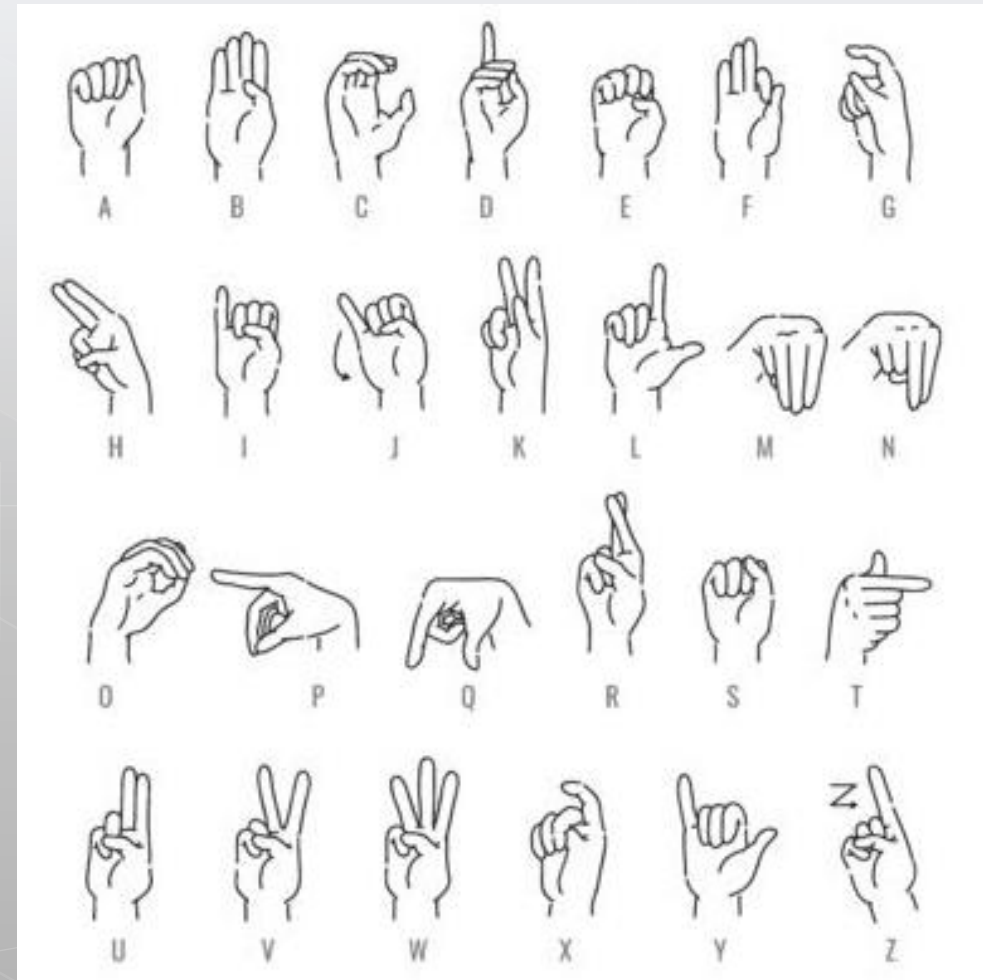
Atoany Nazareth Fierro Radilla.



INTRODUCCIÓN

La interacción computadora-humana es un área de investigación multidisciplinaria que tiene diferentes aplicaciones dentro de las cuales se encuentra el reconocimiento de lengua de señas.

En Estados Unidos hay cerca de dos millones de personas que viven con discapacidad de origen auditivo, es aquí donde cobra relevancia la Lengua de Señas Americana o por sus siglas en inglés (ASL) y su relación con las tecnologías recientes.



La Lengua de Señas Americana (ASL).

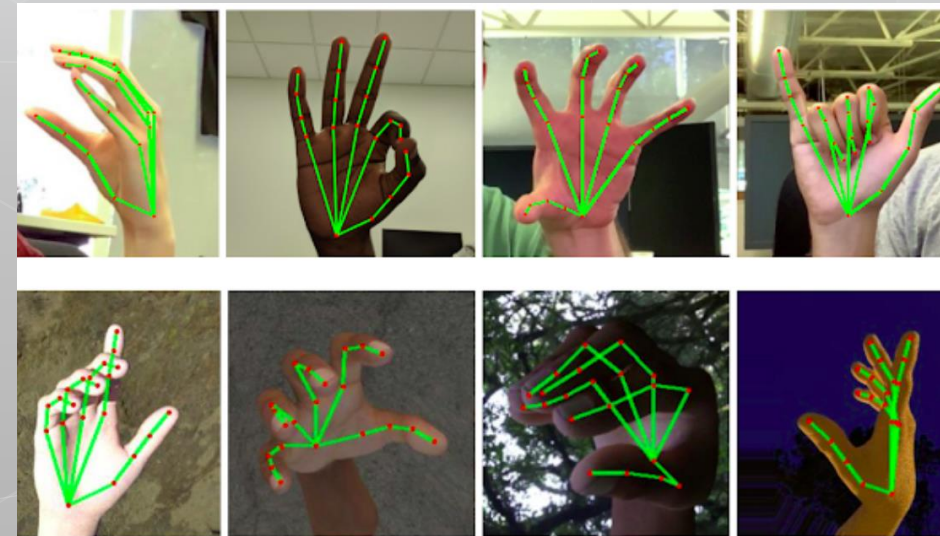
INTRODUCCIÓN

Las diferentes soluciones que existen hoy en día para el reconocimiento de la Lengua de Señas se dividen en dos áreas:

- Los métodos basados en sensores.
- Los métodos basados en visión por computadora.



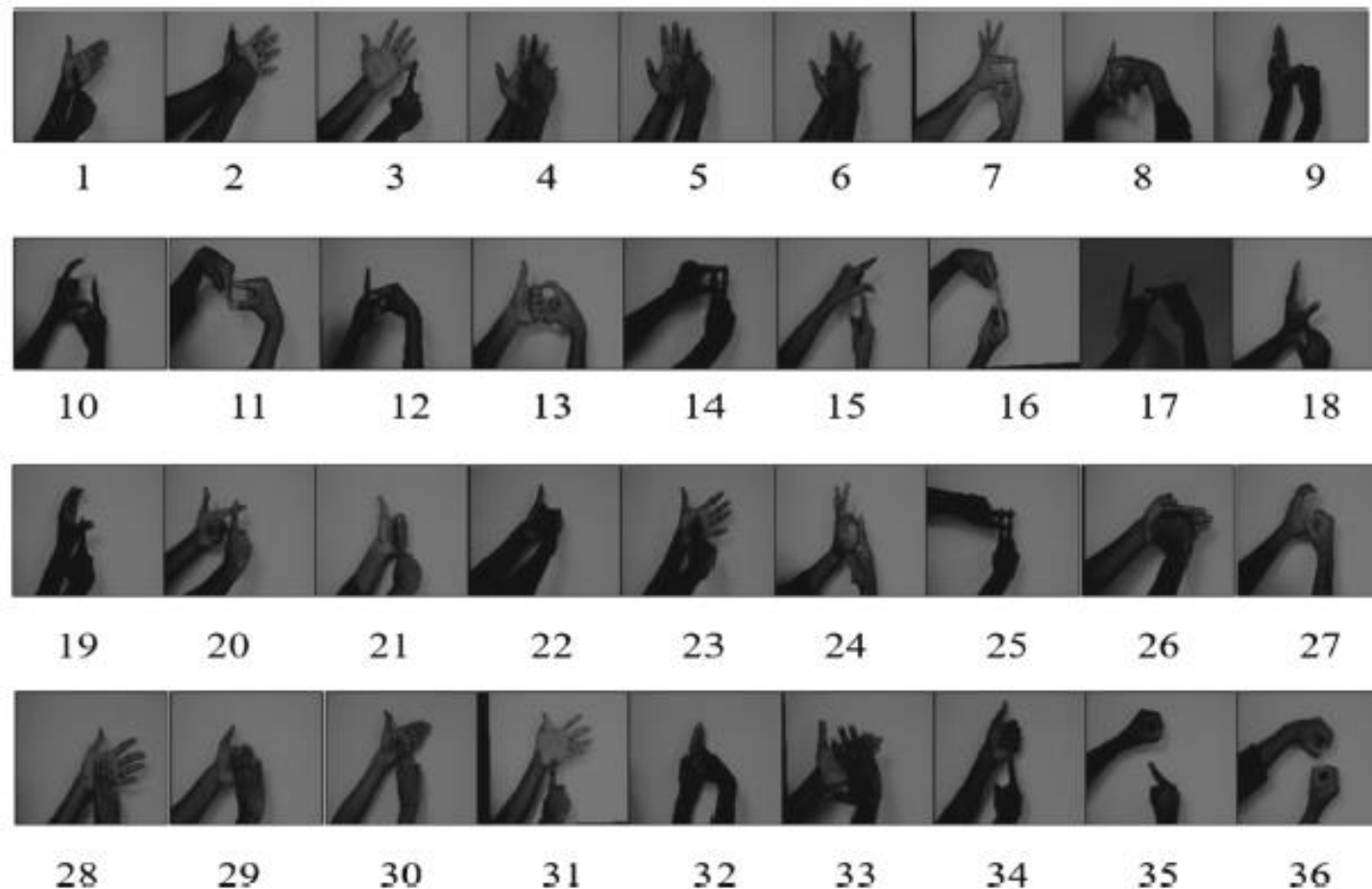
Reconocimiento de la Lengua de Señas basado en sensores.



La visión por computadora y su implicación en la Lengua de Señas.

INTRODUCCIÓN

Sin embargo, dentro del área de la visión por computadora, el deletreo manual en un sistema básico basado en colores sufre problemas al momento de segmentar las señas de la mano tales como antecedentes complejos y variaciones entre correlaciones entre clases e intraclases.

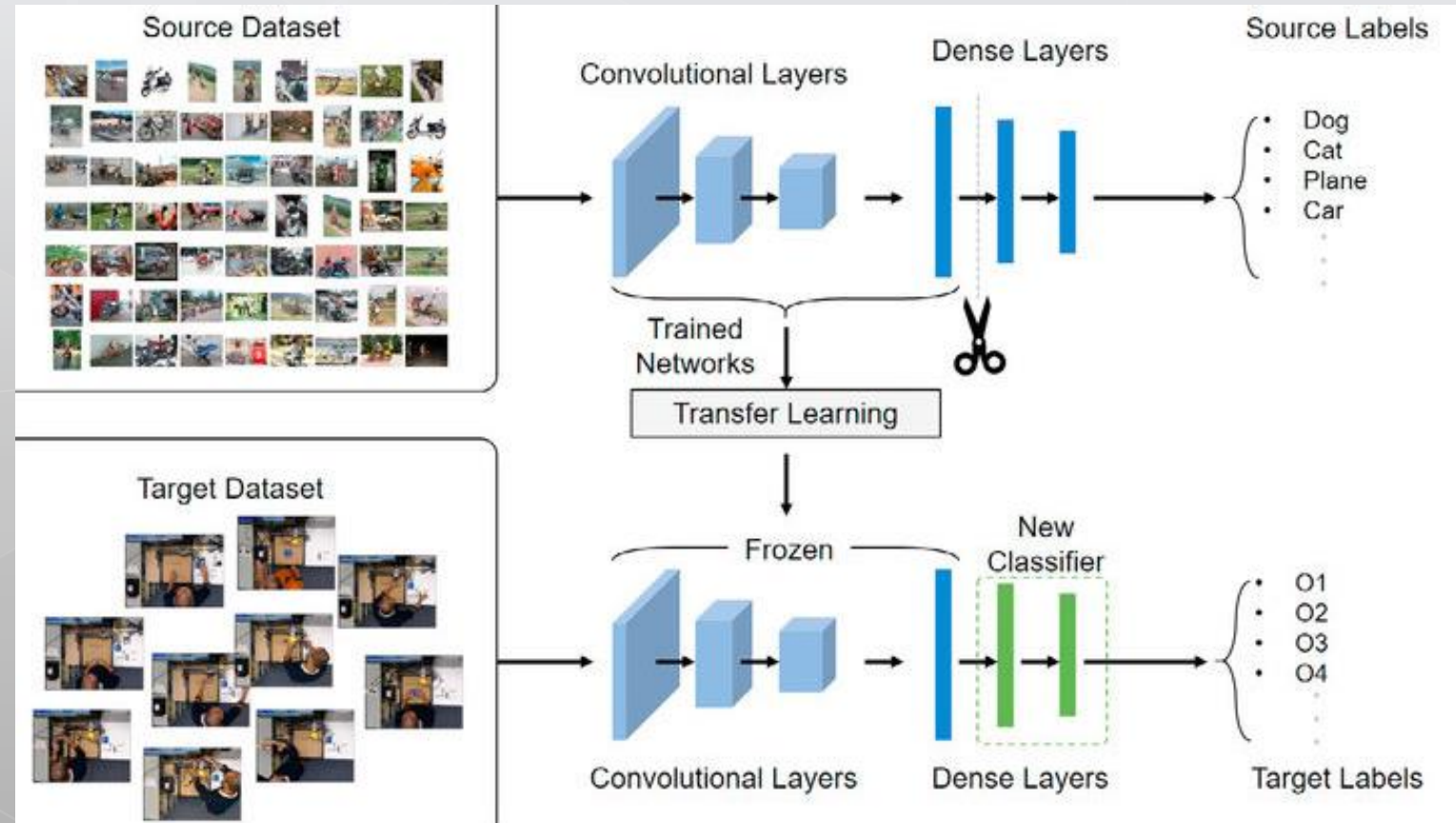


La visión por computadora en combinación con un algoritmo de detección de patrones es la base de este método.

METODOLOGÍA

Todo ello impulsa el desarrollo de nuevos sistemas de reconocimiento de Lengua de Señas utilizando redes neuronales, la creación de una arquitectura de red neuronal que acelere el aprendizaje por medio del aprendizaje transferido.

El aprendizaje de transferencia es una técnica basada en redes preentrenadas, se elimina la última parte de la red (el clasificador) y la es reemplazada con el objetivo. Posteriormente, se entrena al clasificador.

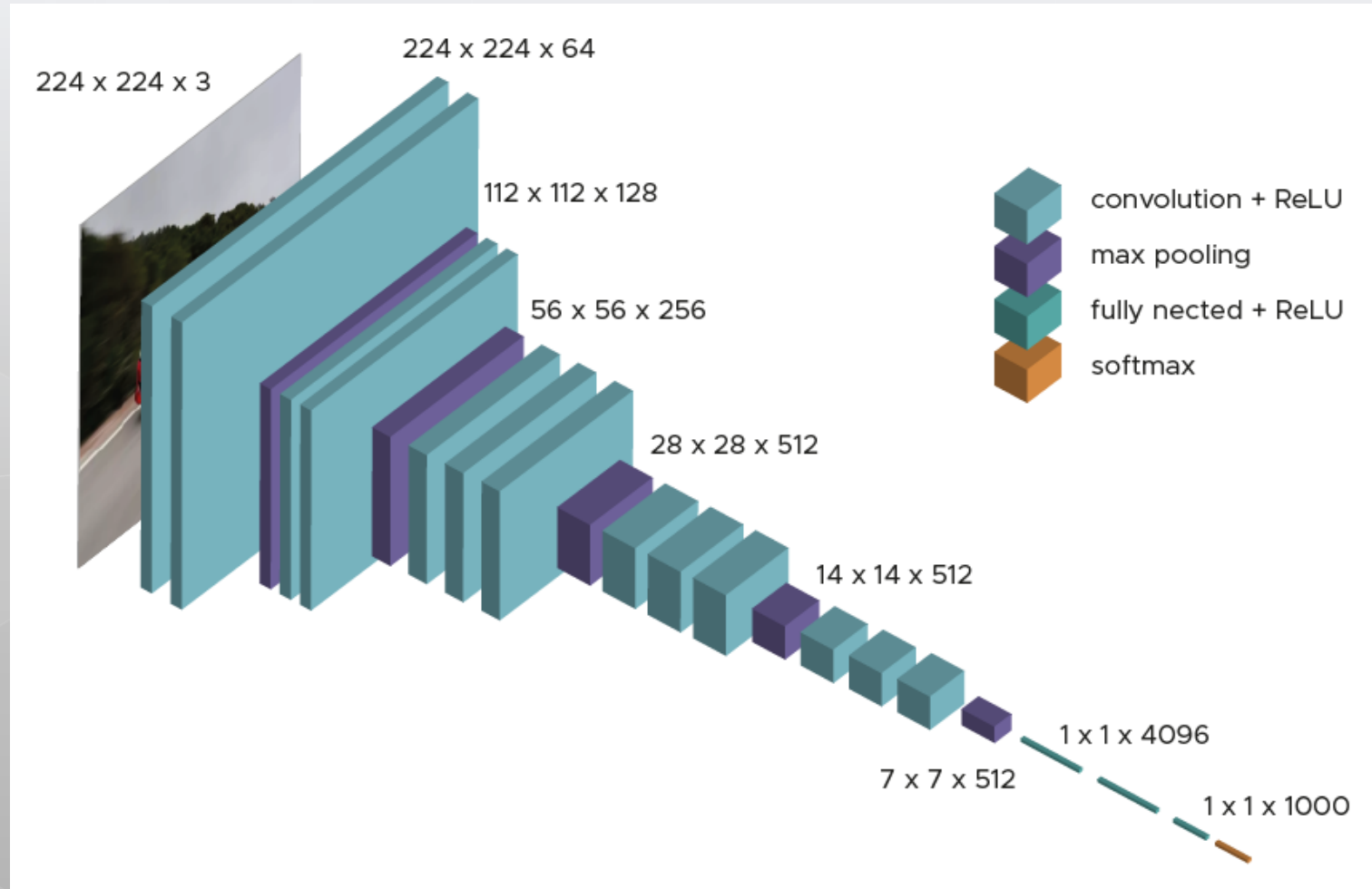


Transfer learning.

METODOLOGÍA

Por otro lado, involucrar a VGG16 es imprescindible.

VGG16 es una red compuesta por 16 capas y que fue entrenada con la base de datos ImageNet, basado en reemplazar los filtros de los kernel por un conjunto de filtros de tamaño 3×3 .



La red VGG16.

METODOLOGÍA

Las bases de datos utilizadas son públicas y de origen descentralizado:

- *Training Dataset*: 27455 imágenes.
- *Test Dataset*: 7172 imágenes.

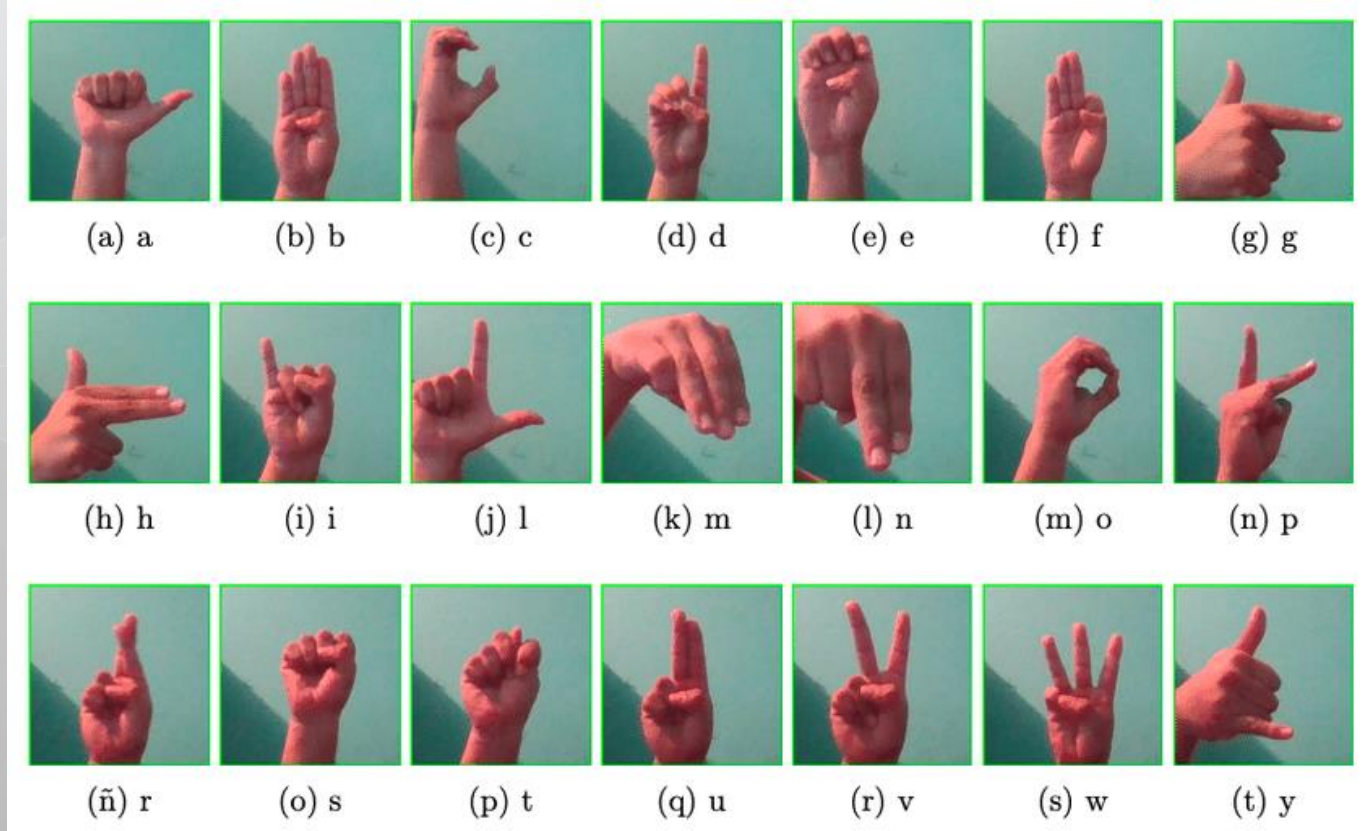
Arquitectura: VGG16

Épocas del proceso: 20

Software:

- Tensorflow.
- Keras.

Tamaño de las imágenes: 280*280px.



Existe una alta disponibilidad pública de volúmenes de datos relacionados a la Lengua de Señas.

RESULTADOS Y ANÁLISIS

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
0	0	0	0	0	13	2	0	0	0	4	0	115	104	10	69	1	0	12	0	0	0	0	1	0
1	0	150	17	5	0	0	5	0	0	133	0	3	50	0	16	1	5	0	0	1	0	45	1	0
2	0	0	204	0	0	3	0	0	0	1	0	0	0	51	42	0	0	4	0	0	0	0	5	0
3	0	0	0	4	0	3	21	0	0	55	37	0	29	0	26	0	29	0	0	0	41	0	0	0
4	0	0	0	0	70	0	0	0	0	67	0	51	58	0	206	2	0	44	0	0	0	0	0	0
5	0	0	3	0	0	92	0	0	0	11	0	11	20	0	82	0	0	5	0	0	19	4	0	0
6	0	0	0	0	0	0	93	0	0	20	0	52	3	0	145	1	0	0	0	0	0	0	2	32
7	0	0	0	0	0	0	54	13	0	3	0	10	12	32	208	44	0	1	0	0	0	0	0	59
8	0	0	0	1	0	0	16	0	4	49	17	15	67	4	105	5	0	4	0	0	1	0	0	0
9	0	0	0	0	0	0	22	0	0	250	0	0	0	0	38	0	5	0	0	0	11	1	0	4
10	0	0	1	8	0	0	49	0	49	2	78	0	0	0	22	0	0	0	0	0	0	0	0	0
11	0	0	0	0	10	0	8	0	0	32	0	72	69	1	198	0	0	4	0	0	0	0	0	0
12	0	1	0	0	29	0	0	0	0	16	0	4	51	0	170	5	0	14	0	0	0	1	0	0
13	0	0	2	0	6	16	3	0	0	17	0	24	17	124	37	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	14	0	0	0	1	332	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	2	0	0	13	0	3	21	1	90	34	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	12	0	0	93	0	0	14	0	0	0	24	0	0	0	1	0	0	0
17	0	0	0	0	1	3	12	0	0	5	0	27	94	0	66	0	0	35	0	3	0	0	0	0
18	0	0	28	0	0	0	22	0	0	57	0	0	1	0	62	0	0	0	36	0	0	0	42	0
19	0	1	0	26	0	0	14	0	0	112	1	0	41	0	18	0	18	0	0	20	5	10	0	0
20	0	0	0	0	0	0	25	0	0	175	0	6	37	0	17	0	2	0	0	0	80	3	0	1
21	0	0	0	0	0	0	8	0	0	94	0	0	31	0	13	0	2	0	0	0	3	55	0	0
22	0	0	19	0	0	12	3	0	0	85	0	0	0	4	61	2	0	10	0	0	0	0	71	0
23	0	0	0	0	0	0	5	0	0	18	0	0	12	0	185	4	0	0	0	0	1	0	0	107

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
0	0	0	0	2	0	4	0	0	0	1	0	31	26	0	29	0	0	19	0	0	0	0	0	0
1	0	27	1	2	1	2	6	0	0	12	0	0	37	0	3	0	1	0	0	1	0	8	0	0
2	0	0	90	0	0	0	0	0	0	0	0	0	1	2	16	3	0	0	0	0	0	0	2	0
3	0	0	0	23	0	6	20	0	0	35	2	0	5	0	18	0	2	1	0	1	5	0	0	1
4	0	0	1	0	34	0	0	0	0	3	0	6	19	0	24	0	0	8	0	0	0	0	0	0
5	0	0	4	0	0	27	1	0	0	9	0	1	9	0	56	0	0	5	0	0	6	1	1	0
6	0	0	3	0	0	1	36	0	0	1	0	12	0	0	50	1	0	0	0	0	0	0	1	4
7	0	0	0	0	0	0	17	12	0	1	0	10	9	2	39	1	0	0	3	0	0	0	0	7
8	0	5	1	0	0	2	3	0	0	16	0	4	19	1	43	0	0	10	0	2	4	3	1	2
9	0	0	0	0	2	0	14	0	0	81	0	0	0	0	7	0	3	0	0	0	4	0	0	0
10	0	0	1	1	0	0	19	0	7	1	80	0	6	0	4	0	0	0	3	0	1	0	0	1
11	0	0	0	0	0	0	2	0	0	9	0	18	14	0	59	0	0	3	0	0	0	0	0	0
12	0	0	0	0	7	3	0	0	0	9	0	0	26	0	68	0	0	2	0	0	0	0	0	0
13	0	0	10	0	0	0	0	0	0	4	0	5	11	72	12	0	0	5	0	0	0	0	0	0
14	0	0	2	0	0	0	1	0	0	1	0	0	0	0	104	0	0	0	0	0	0	0	0	0
15	0	0	2	0	0	3	5	0	0	6	0	2	26	2	58	23	0	0	0	0	0	0	0	0
16	0	0	0	4	0	0	12	0	0	46	0	0	2	0	8	1	44	0	0	0	10	2	0	0
17	0	1	0	0	2	0	1	0	0	17	0	11	26	0	50	0	0	11	0	0	0	0	0	0
18	0	0	2	1	0	0	20	0	0	21	0	3	9	2	37	0	0	0	16	3	0	1	3	0
19	0	0	1	0	0	2	7	0	0	50	1	0	13	0	10	0	14	5	0	1	10	2	0	0
20	0	0	2	1	0	1	12	0	0	26	0	0	13	0	9	1	3	0	0	0	35	5	0	0
21	0	0	0	0	0	0	0	0	0	41	0	0	21	0	6	2	3	0	0	0	3	46	0	0
22	0	0	8	0	0	2	3	0	0	32	0	1	4	1	34	0	0	5	0	0	0	2	24	0
23	0	0	0	0	0	0	9	0	0	2	0	0	5	1	55	0	0	0	0	0	1	0	0	38

Primera y segunda matrices de confusión, el antes y el después del aprendizaje transferido a la red neuronal.

CONCLUSIONES

Como es posible observar en los experimentos y en las matrices de confusión, el proceso de transferencia de aprendizaje es capaz de detectar el alfabeto reunido por la lengua de señas con una mayor precisión en los resultados.

Este modelo partió desde el entrenamiento inicial que, al truncar las capas de características generales y al entrenar las capas de características específicas, denominadas capas superiores del modelo, se evitó la aparición de *overfitting*.

De esta manera, se obtienen resultados más precisos, en cuanto a la clasificación del alfabeto de la lengua de señas, tal y como se muestra en la experimentación del reporte de clasificación final, se tuvo una precisión del 32%, una mejor cifra que la disponible después de diseñar el modelo desde el inicio, a diferencia del primer caso de experimentación que obtuvo una precisión del 10%.

Fuentes de información

- Chapin, W., Kramer, J., Haas, C., Leifer, L., Macken, E.: TeleSign: A Sign Language Telecomm- unication System. IEEE (1992).
- D. Warchoř, T. Kapu'sciński, and M. Wysocki, "Recognition of fingerspelling sequences in polish sign language using point clouds obtained from depth images," Sensors, vol. 19, no. 5, p. 1078, 2019.
- J. Singha, A. Roy, and R. H. Laskar, "Dynamic hand gesture recognition, 2016 M. Mohandes, M. Deriche, and J. Liu, "Image-based and sensor-based approaches to arabic sign language recognition," IEEE transactions on human-machine systems, vol. 44, no. 4, pp. 551–557, 2014.
- National Center of Birth Defects and Developmental Disabilities home page, <http://www.cdc.gov/ncbddd/bd/bdres.ht>
Bitton, G. (1994) Wastewater Microbiology, Wiley-Liss, New York, 478 pp.