

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

EFFECT OF TRANSFER LEARNING IN VGG16 FOR AMERICAN SIGN LANGUAGE ALPHABET RECOGNITION

Regina Alexia Blas Flores¹
Armando Arredondo Valle¹
Emiliano Vivas Rodriguez¹
Atoany Nazareth Fierro Radilla^{1*}

Tópico del congreso: Ingeniería computacional
Subtópico del congreso: Aprendizaje automático

Abstract

Computer-human interaction is a multidisciplinary research area that has different applications, one of which is sign language recognition. Sign language is based on signs and gestures, which are usually classified as non-verbal communication. Despite this, it is possible to transmit the desired message among people with hearing problems. Currently, many recognition technologies for this interaction method remain rudimentary. The learning transfer process with neural networks is a software reengineering process that consists of reusing a deep and truncated neural network in some of its layers to be implemented for a purpose like that of its creation. The application of this process and computer vision in everyday life situations, such as communication and interactions with individuals, would increase people's productivity. Undoubtedly, the involvement of the most innovative computer technologies represents a compromising transition for the social inclusion of people with disabilities.

Key Words: AS, convolutional neural networks, transfer learning, classification, VGG16.

¹ Departamento de Computación, Departamento de Mecatrónica, Tecnológico de Monterrey

* Autor correspondiente: Departamento de Mecatrónica, Tecnológico de Monterrey. México 95D km 104, Real del Puente, Xochitepec, Morelos. 62790. México. Email: afierro@tec.mx

Resumen

La interacción computadora-humana es un área de investigación multidisciplinaria que tiene diferentes aplicaciones, dentro de las cuales se encuentra el reconocimiento de la lengua de señas. La lengua de señas se basa en signos y gestos, que se suelen clasificar como una comunicación no verbal. A pesar de ello, se logra transmitir el mensaje deseado entre personas con problemas de audición. Actualmente, bastantes tecnologías de reconocimiento de este método de interacción continúan siendo rudimentarias. El proceso de transferencia de aprendizaje con redes neurales es un proceso de reingeniería de software que consiste en reutilizar una red neuronal profunda y truncada en alguna de sus capas para implementarse en algún propósito semejante al de su creación. La aplicación de este proceso junto con la visión por computadora en situaciones de la vida cotidiana, como la comunicación e interacciones con individuos, incrementaría la productividad de las personas. Sin duda, la implicación de las tecnologías de la computación más innovadoras y novedosas representan una transición comprometedora para la inclusión social de personas en situación de discapacidad.

Palabras clave: ASL, Redes Neuronales Convolucionales, Transferencia de Aprendizaje, Clasificación, VGG16.

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 (National Center of Birth Defects and Developmental Disabilities et al, 2021).

En Estados Unidos hay cerca de dos millones de personas que viven con discapacidad de origen auditivo y la Lengua de Señas Americano o por sus siglas en inglés (ASL) (Chaping et al., 1992), es el principal método de comunicación para ellos. Sin embargo, no todos pueden comprender esta lengua por lo tanto se propone un modo de traducir la Lengua de Señas para las demás personas.

La lengua de Señas se basa en signos y gestos, que se suelen clasificar como una comunicación no verbal. A pesar de ello, se logra transmitir el mensaje deseado entre personas con problemas de audición. (Warchoř et al., 2019). Las diferentes soluciones que existen hoy en día para el reconocimiento de la lengua de señas se basan en sensores y en métodos de visión (Mohandes et al., 2014). Las soluciones con sensores suelen utilizar guantes especiales o un sensor de presencia que pueda determinar la orientación de la mano y los diferentes movimientos que se estén realizando (Singha et al., 2016). Por otro lado, el principio de los métodos de visión es por medio de una cámara que sea capaz de capturar las señas sin necesidad de un sensor, para este método se emplean diversos procesadores de imágenes y se utilizan técnicas de aprendizaje automático con la finalidad de analizar y representar las señas. Recientemente, el que existan cámaras con precios cada vez más accesibles, ha permitido que los investigadores las utilicen en muchas aplicaciones de visión.

Estos sistemas siempre se basan en el color para capturar la información de la textura y formas de los gestos de las manos. El sistema de reconocimiento automático de lengua de señas puede ser clasificarse en las siguientes clases, interpretación de oraciones, palabras y reconocimiento por el alfabeto de manera manual. El sistema de reconocimiento del alfabeto de la lengua de señas es considerado como una parte esencial para el aprendizaje para nuevos usuarios. Estos sistemas siempre se basan en el color para capturar la información de la textura y formas de los gestos de las manos. 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.

Sin embargo, los métodos basados en profundidad de capas superan dichos problemas, ya que descartan aquella información que no es relevante mientras seleccionan y utilizan la información mandada por distancia de la cámara de aquella que cumpla con las especificaciones indicadas.

En este trabajo se propone un nuevo sistema de reconocimiento de lengua de señas utilizando redes neuronales, manejando una arquitectura de red neuronal que acelere el aprendizaje, el aprendizaje transferido consiste en aprovechar, la mayor cantidad de información relacionada con la solución de un problema y posteriormente ser usada sobre otro, pero ambos deben compartir ciertas características. (El transfer learning y las redes convolucionales, 2020).

El proceso de transferencia de aprendizaje con redes neurales

El aprendizaje y el entrenamiento constante de la computadora nos permiten obtener resultados, tales como una correlación cercana, así como alta al momento de realizar pruebas como se muestra el en el ejemplo.

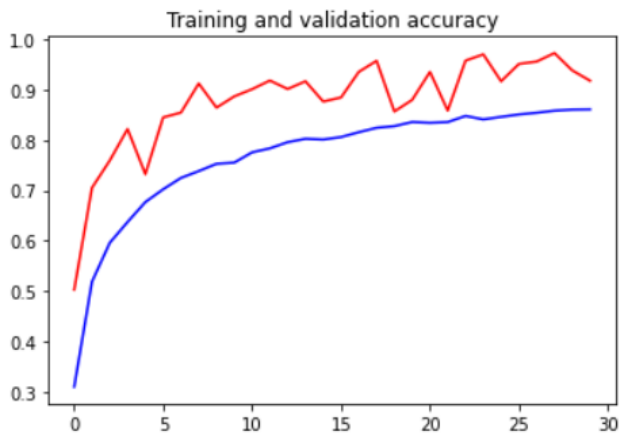


Figura 1. Presición de la prueba y la validación.

Justificando de esta forma el uso de nuestras muestras de entrenamiento, así como de validación. También sustenta esto mismo el índice de pérdidas obtenido.



Figura 2. Pérdida de la prueba y la validación.

Sin embargo, este mismo índice de pérdidas cobra factura y es mejor verlo en la matriz de confusión, donde se tienen más flujos de información, siendo alarmante como ya lo avisa la gráfica anterior los valores 4,5,8 y 20 donde peor clasificación tiene utilizando las muestras de entrenamiento.

	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

Figura 3. Primera matriz de confusión.

De igual manera, en la validación, es donde se aprecia un cambio ligero en las alarmantes cifras que si bien no son perfectas se desplazan a otros valores como lo son el 7 con el 14, el 9 con el 20 y así sucesivamente.

Siendo, incluso, posible observar esto en la consola, teniendo valores de score en f1 bajos.

	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	2	0	4	0	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

Figura 4. Segunda matriz de confusión.

Conclusiones

El proceso de transferencia de aprendizaje con redes neurales es un proceso de reingeniería de software que consiste en reutilizar una red neuronal profunda y truncada en alguna de sus capas para implementarse en algún propósito semejante al de su creación. Después de hacer un entrenamiento liviano en algún contexto ajeno a sus inicios, pero similar a ellos, es posible obtener un modelo matemático con un mejor desempeño y menor coste computacional previo y durante la ejecución del algoritmo. Como se puede observar en los experimentos, el proceso de transferencia de aprendizaje es capaz de detectar el alfabeto reunido por la lengua de señas.

Cabe resaltar que, 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* en la base de datos existente. 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. Finalmente, este reporte obtuvo 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%.

Agradecimientos

Los autores agradecen al TEC de Monterrey campus Cuernavaca por las facilidades proporcionadas para el desarrollo de este artículo, así como también al resto de nuestros compañeros, Ana Laura García Carreño, Diane Zenil Lagunas, Mariana Dirzo Sánchez y Joana Itzel Barreto López del grupo estudiantil Signers.AI del campus.

Referencias bibliográficas

- Chapin, W., Kramer, J., Haas, C., Leifer, L., Macken, E.: TeleSign: A Sign Language Telecommunication 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.htm>
- Bitton, G. (1994) *Wastewater Microbiology*, Wiley-Liss, New York, 478 pp.
- El transfer learning y las redes convolucionales*.2020. Viewnext. Recuperado 23 de septiembre de 2022, de <https://www.viewnext.com/transfer-learning-y-redes-convolucionales/>
- COGERH, Companhia de Gestão dos Recursos Hídricos (2015) *Monitoramento Quantitativo e Qualitativo dos Recursos Hídricos, Campanha Novembro/2014*. Acesso em 03 de fevereiro de 2015, disponível em: <http://www.hidro.ce.gov.br/reservatorios/qualidade/eutrofizacao>
- García, J. y Romero, H. (1997) El tratamiento de las aguas residuales de ciudades medianas: problemática y alternativas de solución, en *Memorias Técnicas XI Congreso Nacional de Ingeniería Sanitaria y Ciencias Ambientales*, Tomo1, FEMISCA A.C., 4 a 7 de noviembre, Zacatecas Zac. 243-250.