HOJA DE ESTILO PARA COMUNICACIONES Y PÓSTERS

INVESTIGACIÓN

Tipo de aportación: *Marque con una X el tipo de aportación que presenta.*

|  |  |
| --- | --- |
| X | Comunicación - Investigación |
|  | Póster - Investigación |

Temática de la aportación: *Marque con una X la temática de la aportación que realiza.*

|  |  |
| --- | --- |
|  | Aprendizaje y desarrollo profesional en la Sociedad 5.0 |
|  | Gobernanza de Instituciones en la Sociedad 5.0 |
|  | Herramientas y habilidades digitales en las Organizaciones |
|  | Inteligencia artificial generativa: un aliado ante la transformación |
|  | Convergencia entre la organización formal e informal en las Organizaciones |
| X | Convergencia entre el mundo físico y el digital en las Organizaciones |
|  | Liderazgos necesarios para la transformación |
|  | Gestión del cambio y autonomía: personal y organizacional |
|  | Ética y responsabilidad digital |
|  | Experiencias transformadoras: Robótica educativa, Robots sociales, Realidad Virtual, Realidad aumentada, Simulaciones, Herramientas digitales para el STEAM,… |

**APLICACIÓN DE REDES TRANSFORMER PARA DISCRIMINACIÓN DE PRODUCTOS LÍQUIDOS DE ASEO**

**Robinson Jiménez-Moreno, Anny Espitia-Cubillos, Esperanza Rodríguez-Carmona**

Universidad Militar Nueva Granada- Grupo de investigación DAVINCI /Colombia

***Resumen***

El presente documento presenta un algoritmo de inteligencia artificial basado en redes neuronales transformer que permite la discriminación de productos líquidos de aseo identificados con distintas etiquetas y con presentaciones en varios colores, desde una cámara para facilitar la gestión de su manipulación posterior mediante un computador lo que permite hacer realidad, en entornos manufactureros, la conexión entre el mundo físico y el digital. El proceso parte de la digitalización de los productos lo que permite establecer una base de datos, posteriormente se definen los parámetros para el entrenamiento de la red, que una vez llevado a cabo es evaluado mediante la medición del tiempo de aprendizaje, la precisión y el tiempo que tarda en la clasificación de los productos, todo ello desarrollado en un ambiente virtual. Los resultados permiten concluir que incluso con una base de datos reducida que incluye visualizaciones incompletas y borrosas de las etiquetes, los resultados de precisión en la clasificación son del 100%, sin ningún tipo de confusión entre categorías, con tiempos de procesamiento inferiores a 0.5 segundos, dada la robustez de la red transformer seleccionada.

**APLICACIÓN DE REDES TRANSFORMER PARA DISCRIMINACIÓN DE PRODUCTOS LÍQUIDOS DE ASEO**

**Robinson Jiménez-Moreno, Anny Espitia-Cubillos, Esperanza Rodríguez-Carmona**

Universidad Militar Nueva Granada- Grupo de investigación DAVINCI /Colombia

***1. Introducción***

La industria 4.0 es un concepto implementado a distintos niveles en las empresas manufactureras (Habib, Bnouachir, Chergui, & Ammoumou, 2022), algunas aplicaciones se fundamentan en el uso de la inteligencia artificial, usando el modelo Faster RCNN (Saleem, Potgieter, & Arif, 2022), sistemas de aprendizaje profundo (Vaddadi, y otros, 2022) (Wu, 2023), (Yoon, Han, & Nguyen, 2023), aplicando IoT (Bose, Mondal, Sarkar, & Roy, 2022) o con arquitecturas como la YOLO (Qureshi, y otros, 2024) (Qi & Sun, 2024) (Zheng, Chen, Cheng, Du, & Jiang, 2024).

(Mark, Rauch, & Matt, 2022) proponen entender la inteligencia artificial como un instrumento que facilita la ejecución de actividades y soporta a los tomadores de decisiones, visión compartida en el presente desarrollo.

Gracias a la estandarización de los productos, una de las actividades susceptible de automatización es su identificación, lo que permite la medición en tiempo real de la productividad, los niveles de inventario de cada producto, su ubicación, entre otras, facilitando la toma oportuna de decisiones informadas en los procesos productivos y logísticos. Las redes transformer (Touvron, Cord, El-Nouby, Verbeek, & Jégou, 2022) sirven a este propósito y han tenido aplicaciones a nivel industrial previas exitosas que se exponen a continuación.

(Ma, y otros, 2024) usaron una red transformer jerárquica para reconocer operaciones de producción industrial de ropa con una agrupación de redes para mejorar la precisión y reducir la sobrecarga.

Para procesos químicos, (Wang, Sun, Wang, & Luo, 2023), proponen un modelo de predicción basado en una red transformer mejorada. Para la optimización de los procesos del petróleo (Huang, y otros, 2024) plantean un modelo llamado Time Patch Dynamic Attention Transformer (TPDAT) que segmenta los datos para mejorar el reconocimiento de sucesos y fluctuaciones transitorias; (Ma, Li, & Yuan, 2024) emplean un modelo Transformer para procesar los datos de operación de un pozo en pro de mejorar la eficiencia en la producción, reducción de costos de mantenimiento e incrementar la vida útil de los equipos.

Para pronosticar la vida útil de los motores de aeronaves se emplea una red Transformer, que se compone de un mecanismo que extrae y pondera características de forma dinámica obteniendo un buen desempeño en detección de objetos, previsión del tráfico, segmentación de imágenes, riesgo de fallas y reducción de pérdidas (Liu, Song, & Zhou, 2022). En el campo energético (Al-Ali, y otros, 2023) con un modelo CNN-LSTM-Transformer (Convolutional Neural Network (CNN), a Long Short-Term Memory (LSTM) network and a Transformer) pronostican la producción de energía solar usando clústeres para analizar la correlación de los datos de entrada junto con sus características dinámicas, los resultados tuvieron mayor precisión que los obtenidos con modelos como el LSTM-CNN.

***2. Metodología***

El proceso parte de la digitalización de los productos de interés para establecer una base de datos de imágenes para el entrenamiento de una arquitectura de red por transferencia de aprendizaje basada en modelos transformer (Dosovitskiy, y otros, 2021). Posteriormente se definen los parámetros para el entrenamiento de la red, que es evaluado mediante una matriz de confusión, la precisión y el tiempo que tarda en la clasificación de los productos.

La base de datos incluye siete de los productos de más demanda en una empresa de fabricación de insumos de limpieza, que son envasados en el mismo tipo de recipiente y que en algunos casos tiene el mismo color. La diferenciación especifica entre cada uno corresponde a la etiqueta de producto adherida al recipiente, la base de datos incluye imágenes de la vista frontal de la etiqueta y vistas laterales que permitan desde el 50% de su visibilidad. De igual forma la resolución de las imágenes varia en relación a la distancia focal de la cámara, en algunos casos el texto de la etiqueta no es claramente legible, dificultando la labor de reconocimiento, por lo cual arquitecturas convencionales no han dado buen resultado, requiriendo por ejemplo integración de técnicas como CNN y SURF (Guacheta-Alba, Espitia-Cubillos, & Jiménez-Moreno, 2024). La base de datos es reducida, dado que las características de identificación de productos se centran en el envase, tomando de cada producto 20 imágenes y a su vez dividiéndolas en 75% para entrenamiento (15 imágenes), 10% para validación (2 imágenes) y 15% para prueba (3 imágenes), por categoría. Una muestra de la base de datos se puede evidenciar en la tabla 1.

**Tabla 1.** Categorías y muestras de la base de datos empleada.

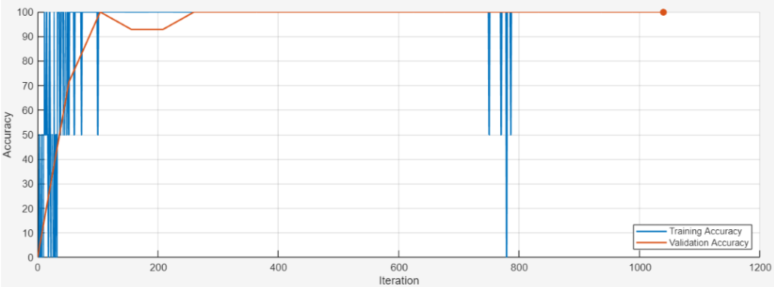
|  |  |
| --- | --- |
| **Producto** | **Ejemplos de entrenamiento** |
| Blanqueador | Botella de plástico  Descripción generada automáticamente con confianza baja |
| Bluemax | Interfaz de usuario gráfica, Aplicación  Descripción generada automáticamente |
| Desengrasante | Botella de plástico  Descripción generada automáticamente con confianza baja |
| Detergente | Botella de plástico  Descripción generada automáticamente con confianza baja |
| Lavaloza | Botella de plástico  Descripción generada automáticamente con confianza media |
| Limpiador | Icono  Descripción generada automáticamente con confianza media |
| Suavizante | Botella al lado de un refrigerador  Descripción generada automáticamente con confianza baja |

La tabla 2 ilustra los parámetros de entrenamiento definidos, que derivado de la base de datos requiere un mini Batch-size pequeño, para el caso de 2. El volumen de entrada de la red requirió un redimensionamiento previo de las imágenes a un tamaño de 348 pixeles por cada lado, como imágenes a color (RGB), implicando un volumen de profundidad 3 por cada componente de color. Dada la arquitectura predefinida de la red que cuenta con 143 capas de profundidad, siendo bastante robusta, se emplean pocas épocas de entrenamiento, para el caso 20, obteniendo que para la época 5 ya se alcanza una exactitud del 100%, como se aprecia en la figura 1.

**Tabla 2.** Parámetros de entrenamiento de red.

|  |  |
| --- | --- |
| **Parámetro** | **Valor** |
| Volumen de entrada | 384X384X3 |
| Minibatchsize | 2 |
| Optimizador | SGDM |
| Épocas | 20 |
| Rata de aprendizaje | 0.0001 |
| Tiempo de entrenamiento | 8 MIN 55 SEG |
| Numero de clases | 7 |

**Figura 1.** Progreso en el entrenamiento de la red.



***3. Resultados***

Mediante el análisis por matriz de confusión se logra evidenciar que el nivel de exactitud alcanzado correspondiente al 100% se ve reflejado en el conjunto de imágenes de prueba (3 por categoría). La figura 2 ilustra que no hay confusión de clases, lo cual para el bajo número de imágenes empleado destaca la alta capacidad de la red en la identificación de cada clase con las condiciones de aprendizaje que exige la base de datos.

**Figura 2.** Matriz de confusión de prueba de la red

Escala de tiempo

Descripción generada automáticamente

Adicionalmente se emplean otras imágenes fuera del conjunto inicial para testeo final y medición del tiempo de respuesta en la clasificación de la red. Este resultado se aprecia en la tabla 3, donde se destaca cambios de color en el fondo, de ubicación del producto y del tamaño, obteniendo una identificación clara de cada clase con tiempos promedio de 0,467 segundos. Lo cual resalta que, incluso con una arquitectura muy profunda como la empleada, el tiempo de respuesta es suficiente para un trabajo de manipulación de producto por medios automáticos en líneas de producción, sea para despacho, inventario o almacenamiento.

**Tabla 3.** Resultados y tiempos de validación.

|  |  |
| --- | --- |
| Validación 1 | Validación 2 |
| Interfaz de usuario gráfica, Icono  Descripción generada automáticamente | Imagen que contiene botella, cocina  Descripción generada automáticamente |
| 0.442301 segundos. | 0.492344 segundos. |

***4. Conclusiones y Discusión***

Los hallazgos muestran que para caso de arquitecturas transformer robustas, no se requiere una base de datos extensa incluso con datos de entrada de difícil clasificación como los empleados.

Los resultados obtenidos son contrastados con trabajos previos (Guacheta-Alba, Espitia-Cubillos, & Jiménez-Moreno, 2024) que permiten la clasificación de los productos con inteligencia artificial requiriendo modelos híbridos para el reconocimiento de la base de datos similar a la expuesta, de donde se destaca el alto impacto que tiene el uso de redes transformer en aplicaciones orientadas a industria, con tiempo suficientes para manipulación del producto identificado.

Al respecto, (Espitia Cubillos, Jiménez Moreno, & Rodríguez Carmona, 2025) señalan que para seleccionar el modelo más conveniente según los objetivos del usuario final en un entorno de producción industrial los tiempos de entrenamiento y, para el caso, la profundidad de la red no son relevantes, como sí lo son criterios como la precisión y tiempos de clasificación.

***5. Agradecimientos***

Producto derivado del proyecto de investigación titulado “Diseño de un modelo de interacción humano robot mediante algoritmos de aprendizaje profundo”, código INV-ING-3971 financiado por la vicerrectoría de investigaciones de la Universidad Militar Nueva Granada, vigencia 2024.

**REFERENCIAS**

Al-Ali, M., Hajji, Y., Said, Y., Hleili, M., Alanzi, A., Laatar, A., & Atri, M. (2023). Solar Energy Production Forecasting Based on a Hybrid CNN-LSTM-Transformer Model. *Mathematics, 11*(3), 676. doi:10.3390/math11030676

Bose, R., Mondal, H., Sarkar, I., & Roy, S. (2022). Design of smart inventory management system for construction sector based on IoT and cloud computing. *e-Prime - Advances in Electrical Engineering, Electronics and Energy, 2*, 100051. doi:10.1016/j.prime.2022.100051

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., & al., M. D. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. doi:10.48550/arXiv.2010.11929

Espitia Cubillos, A. A., Jiménez Moreno, R., & Rodríguez Carmona, E. (2025). Deep learning architectures for location and identification in storage systems. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 592-601. doi:10.11591/ijai.v14.i1.pp592-601

Guacheta-Alba, J. C., Espitia-Cubillos, A. A., & Jiménez-Moreno, R. (2024). Automated Box Classification in a Virtual Industrial Environment Using Machine Vision Algorithms. *12th International Conference on Control, Mechatronics and Automation (ICCMA)* (págs. 305-310). London, UK: IEEE. doi:10.1109/ICCMA63715.2024.10843920

Habib, F. E., Bnouachir, H., Chergui, M., & Ammoumou, A. (2022). Industry 4.0 concepts and implementation challenges: Literature Review. *2022 9th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, (págs. 1-6). Rabat, Morocco. doi:10.1109/WINCOM55661.2022.9966456

Huang, T., Qian, H., Huang, Z., Xu, N., Huang, X., Yin, D., & Wang, B. (2024). A time patch dynamic attention transformer for enhanced well production forecasting in complex oilfield operations. *Energy, 309*, 133186. doi:10.1016/j.energy.2024.133186

Liu, L., Song, X., & Zhou, Z. (2022). Aircraft engine remaining useful life estimation via a double attention-based data-driven architecture. *Reliability Engineering & System Safety, 221*, 108330. doi:10.1016/j.ress.2022.108330

Ma, Y., Li, X., & Yuan, C. (2024). Intelligent prediction of oil well working conditions based on Transformer. *Journal of Physics: Conference Series, 2901*, 012022. doi:10.1088/1742-6596/2901/1/012022

Ma, Y., Wang, X., Yuan, J., Zhang, L., Chen, J., & Fen, K. (2024). Clothing Detection Action Recognition Based on Hierarchical Transformer Networks. *2024 16th International Conference on Communication Software and Networks (ICCSN)*, (págs. 199-206). Ningbo, China. doi:10.1109/ICCSN63464.2024.10793382

Mark, B. G., Rauch, E., & Matt, D. T. (2022). Systematic selection methodology for worker assistance systems in manufacturing. *Computers & Industrial Engineering, 166*, 107982. doi:10.1016/j.cie.2022.107982

Qi, Y., & Sun, H. (2024). Defect Detection of Insulator Based on YOLO Network. *9th International Conference on Electronic Technology and Information Science (ICETIS)*, (págs. 232-235). Hangzhou, China. doi:10.1109/ICETIS61828.2024.10593675

Qureshi, A., Butt, A., Alazeb, A., Mudawi, N., Alonazi, M., Almujally, N., . . . Liu, H. (2024). Semantic Segmentation and YOLO Detector over Aerial Vehicle Images. *Computers, Materials & Continua, 80*(2). doi:10.32604/cmc.2024.052582

Saleem, M. H., Potgieter, J., & Arif, K. M. (2022). Weed Detection by Faster RCNN Model: An Enhanced Anchor Box Approach. *Agronomy, 12*(7), 1580. doi:10.3390/agronomy12071580

Touvron, H., Cord, M., El-Nouby, A., Verbeek, J., & Jégou, H. (2022). Three things everyone should know about vision transformers. En S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, & T. Hassner (Edits.), *Computer Vision–ECCV 2022* (Vol. 13684 , págs. 497-515). Switzerland: Springer Nature. doi:10.1007/978-3-031-20053-3\_29

Vaddadi, S., Srinivas, V., Reddy, N., Girish , H., Rajkiran , D., & Devipriya, A. (2022). Factory Inventory Automation using Industry 4.0 Technologies. *022 IEEE IAS Global Conference on Emerging Technologies (GlobConET)*, (págs. 734-738). Arad, Romania. doi:10.1109/GlobConET53749.2022.9872416

Wang, S., Sun, H., Wang, Y., & Luo, X. (2023). Prediction of key chemical parameters based on improved Transformer. *2023 4th International Conference on Computer Engineering and Application (ICCEA)*, (págs. 855-859). Hangzhou, China. doi:10.1109/ICCEA58433.2023.10135471

Wu, B. (2023). Motion Control Algorithm for Automatic Welding of Complex Intersecting Line Joints Based on Deep Learning. *2023 International Conference on Mechatronics, IoT and Industrial Informatics (ICMIII)*, (págs. 352-356). Melbourne, Australia. doi:10.1109/ICMIII58949.2023.00073

Yoon, J., Han, J., & Nguyen, T. (2023). Logistics box recognition in robotic industrial de-palletising procedure with systematic RGB-D image processing supported by multiple deep learning methods. *Engineering Applications of Artificial Intelligence, 123*(B), 106311. doi:10.1016/j.engappai.2023.106311

Zheng, H., Chen, X., Cheng, H., Du, Y., & Jiang, Z. (2024). MD-YOLO: Surface Defect Detector for Industrial Complex Environments, Optics and Lasers in Engineering. *178*, 108170. doi:10.1016/j.optlaseng.2024.108170