



Mesa 4C Ciudades Sustentables y Habitables

El Departamento de Ingeniería Industrial y Manufactura
del Instituto de Ingeniería y Tecnología
de la Universidad Autónoma de Ciudad Juárez

Presenta:

Avances en la medición de la demanda en el transporte público con inteligencia artificial

Por: M.C. Roberto Contreras Masse,

Dr. Alberto Ochoa-Zezzatti y

Dra. Vianey Torres

Doctorado en Tecnología

En el marco del Coloquio Institucional 2020

27 de noviembre de 2020

Avances en la medición de la demanda en el transporte público con inteligencia artificial

Roberto Contreras Masse, Dr. Alberto Ochoa-Zezzatti, Dra. Vianey Torres

Introducción

El transporte público es un medio de transporte masivo que trae beneficios a la sociedad y al mismo tiempo presenta retos logísticos y administrativos. Existen dos retos importantes y presentes en la industria de la transportación. El primero gira alrededor de la administración de la demanda, adelantarse a la necesidad que tiene el pasajero, proveer el espacio necesario para que transite cómodo y seguro. El segundo es la recaudación. En países donde se paga en efectivo directamente al operador es complicado enterar de forma exacta el importe del pasaje recolectado.

Como opciones que se pueden explorar está el incrementar el personal que va a bordo de cada unidad y que esté en cada estación, lo cual resultaría en un alto incremento de costo de operación y podría mermar la utilidad trasladando este nuevo costo al pasajero. Otra opción que se puede considerar es el cambio de sistema habilitando máquinas dispensadoras en estaciones concentradoras de rutas y troncales, y también disponibles en cada unidad (Translink, 2020). Finalmente, el modelo híbrido con pago electrónico y en efectivo puede ser otra opción (Gobierno de la Ciudad de México, 2020). Este último no se preocupa por los pagos electrónicos, pero sigue siendo una preocupación enterar el pago en efectivo. Sin embargo, cualquiera de estas posibles soluciones no cubre la estimación de la demanda y únicamente se podría conocer el uso del transporte de forma post mortem, i.e. al final del día se podría conocer el número de pasajeros en función del ingreso reportado.

Aunque en algunas opciones mencionadas anteriormente hay tecnología, no se observa el uso de tecnologías emergentes, como es el caso de la Industria 4.0 (I40) y apoyo por medio de inteligencia artificial (AI). Es por eso por lo que este trabajo aborda la medición de la demanda con estas dos piezas disponibles para cualquier industria como parte de la investigación doctoral. El trabajo está organizado en secciones. La siguiente sección

describe los objetivos, preguntas y justificación de la investigación. La tercera hace una descripción del marco teórico-conceptual. La cuarta sección explica el diseño metodológico, para dar paso a la quinta sección para mostrar los resultados y su discusión. Finalmente, la sexta sección aborda las conclusiones de este trabajo.

Objetivos, preguntas y justificación

El objetivo de esta investigación es proponer un modelo de monitoreo de la demanda del servicio público utilizando componentes de la I40 apoyados por la visión por computadora.

Siguiendo este objetivo surgen la pregunta: ¿De qué forma se puede seleccionar un algoritmo o una herramienta existente para poder detectar personas?

La respuesta a esta pregunta permitirá proponer un modelo confiable y probado combinando la visión por computadora con cámaras inteligentes y estimar la demanda para satisfacerla de forma anticipada. Si esto se logra, entonces estaremos ante un modelo que optimiza el manejo de la demanda y permite contabilizar el pasaje que sube al camión y así poder tener precisión de los ingresos por pasaje.

Marco teórico-conceptual

La I40 nace en Alemania por el año de 2011 y fue declarada estrategia nacional en 2013 para después ser adoptada como un término común en la industria de manufactura en 2015. Ese mismo año, el Boston Consulting Group definió nueve pilares que conforman la I40: Ciberseguridad, Cómputo en la nube, Grandes datos (Big Data), Integración de sistemas, Internet de las Cosas Industrial (IIoT), Manufactura aditiva, Realidad aumentada, Robots autónomos y Simulación (Contreras-Masse, Ochoa-Zezzatti, García, & Mejía, 2019). De estos pilares se toman IIoT, Big Data y Simulación.

Para habilitar las cámaras con inteligencia artificial, es necesario lograr identificar las personas en los cuadros del video. Para esto, existen varios métodos para lograr esto, basados en inteligencia artificial dentro de la categoría *deep learning*. Se eligen cinco algoritmos para esta investigación: Yolo3, Tiny Yolo3, HOGCV, y dos aplicaciones comerciales: AWS Rekognition y Azure Computer Vision.

El algoritmo You Only Look Once (YOLO) está inspirado en el ojo humano, donde de un vistazo puede detectar su entorno y su context. YOLO mejora su desempeño utilizando imágenes del mismo tamaño para su entrenamiento de redes convolutivas de capa sencilla (Redmon, Divvala, Girshick, & Farhadi, 2016). YOLOv2 incrementa su complejidad y también su desempeño con una red de 19 capas llamada Darknet-19 (Redmon & Farhadi, YOLO9000: better, faster, stronger, 2017) y YOLO3 mejora la precisión al usar 53 capas en su modelo Darknet-53 (Redmon & Farhadi, Yolov3: An incremental improvement, 2018).

Los histogramas de gradientes orientados (HOG) se usan desde 2006 para detección de objetos por medio de magnitud y orientación al superponer gradientes verticales y horizontales (Nazir, Jan, & Sajjad, 2018). La implementación en Python es ofrecida por la librería OpenCV y se le conoce como HOGCV.

Las librerías comerciales AWS y Azure son candidatas al ser provistas por líderes del mercado, según el cuadrante mágico de Gartner (2020). Su funcionamiento es por medio de interfaces de programación de la aplicación (API) en la que se envía los bytes de la imagen y la API regresa los objetos encontrados y su probabilidad de confianza (Amazon Web Services, 2020; Azure Computer Vision, 2020).

En la literatura también se encuentra que para medir la exactitud o desempeño de un algoritmo de reconocimiento de objetos se usan las métricas de promedio de precisión (AP) (Lin, Goyal, Girshick, He, & Dollár, 2017) y media del promedio de precisión (mAP) (Xiao, et al., 2019; Padilla, Netto, & da Silva, 2020; Redmon & Farhadi, Yolov3: An incremental improvement, 2018). Otros autores han evaluado la Receiver Operating Characteristic (ROC) y el área bajo el promedio de ROC, tomando como indicadores de entrada la exhaustividad y los falsos positivos (Wang, Zheng, Yang, Luo, & Tang, 2019). Un enfoque experimental fue propuesto midiendo aciertos y errores de males detecciones y detecciones fallidas (Kabakus, 2019). También se han propuesto métodos que toman en cuenta la precisión, la exhaustividad, el F1-Score y un porcentaje de mala clasificación (Gao, 2020).

Diseño metodológico

La metodología propuesta para contestar las preguntas de investigación se presenta en la Ilustración 1. Como primer paso se requiere preparar el conjunto de datos. Esto se puede hacer obteniendo imágenes desde internet, solamente que evitamos el uso de procedimientos manuales. En nuestro experimento se obtuvieron 755 imágenes desde Flickr.com por medio de scripts Python de creación propia. Estas imágenes cambiaron de tamaño para que todas fueran de 600 pixels de ancho.

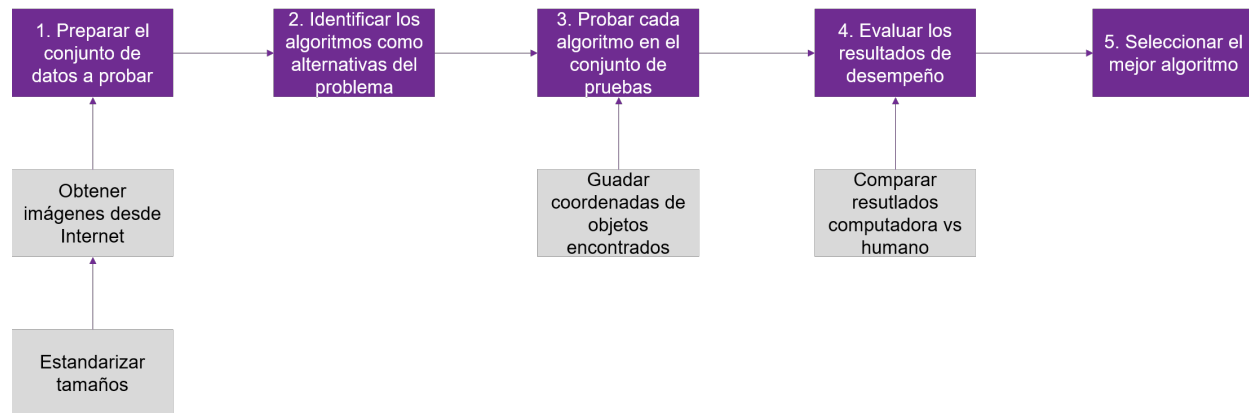


Ilustración 1. Metodología para seleccionar el mejor algoritmo de visión por computadora.

Las alternativas de algoritmos son Yolo3, Yolo3Tiny (usado para dispositivos móviles), HOGCV, AWS Rekognition y Azure. Cada algoritmo se ejecuta con cada una de las fotos del conjunto de datos y se registran los siguientes datos:

- Tiempo de ejecución
- Número de personas encontradas en la imagen
- Tamaño de la imagen (ancho x alto)
- Dimensiones de cada persona encontrada (rectángulo en pixels)
- Probabilidad de confianza
- Se crea la foto analizada con los rectángulos en verde y un punto rojo en el centro del rectángulo de la persona detectada.

Las imágenes originales se deben analizar manualmente para registrar:

- Número de personas en la imagen

Las imágenes producidas por cada algoritmo se deben analizar de forma manual para registrar:

- Personas identificadas correctamente (Positivos)
- Personas identificadas erróneamente (Falsos positivos)
- Personas faltantes de identificar (Falsos negativos)

El análisis para el desempeño del algoritmo implica comparar tiempo de procesamiento para obtener el más rápido, el error entre personas detectadas por el algoritmo y las personas que fueron detectadas manualmente, y el F1-Score mediante la ecuación (1)

$$F1\ Score = 2 \left(\frac{P\ R}{P + R} \right) \quad (1)$$

Donde P es la precisión calculada por $P = TP / (TP + FP)$ y R es la exhaustividad calculada por $R = TP / (TP + FN)$. En estas ecuaciones TP son el número de positivos, FP es el número de falsos positivos y FN es el número de falsos negativos.

Resultados y discusión

Para el análisis de la velocidad de cada algoritmo se observa que Yolo3Tiny es el más rápido, con una media de 0.2086s, seguido por Azure con 0.3846s. Luego AWS con 0.563s, Yolo3 con 1.763s y HOGCV 2.2289s. Al comparar la cantidad de personas detectadas por cada algoritmo, AWS es el que mayor número de personas detecta con una media de 6.0684 personas, Yolo3 con 4.6361, HOGCV con 3.8077, Azure con 2.4981 y finalmente, Yolo3Tiny con 1.9226 personas. No se observa una correlación entre velocidad y personas encontradas, como se muestra en la Ilustración 2.

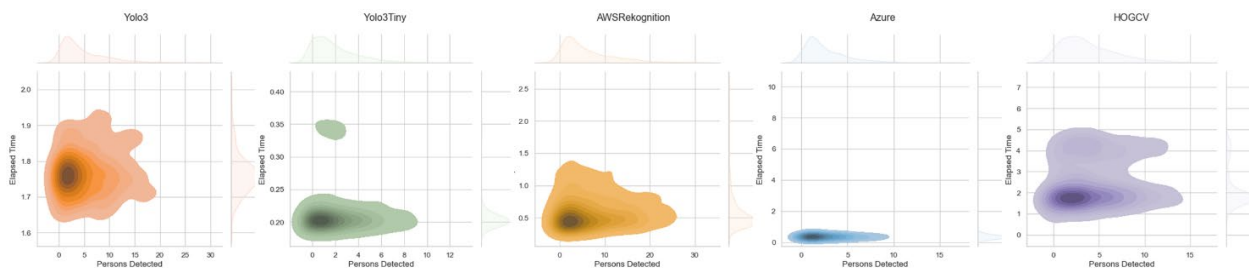


Ilustración 2. Relación tiempo de ejecución vs personas encontradas para todos los algoritmos

Los errores entre lo observado manualmente y lo detectado por los algoritmos no es significativo ni determinante. Si calculamos el error $\varepsilon = x_c - X_h$, donde x_c es el total de personas encontradas por el algoritmo y X_h es el promedio de observaciones manuales, entonces un $\varepsilon < 0$ significa que la computadora detectó menos personas, $\varepsilon = 0$ son aciertos perfectos, y $\varepsilon > 0$ corresponde a los casos cuando el algoritmo detecta más personas que las observadas por el ojo humano. El algoritmo que tuvo más aciertos fue Yolo3 con 225, muy por encima del segundo lugar que fue AWS y Azure, con 160 y 158 aciertos respectivamente. HOGCV y Yolo3Tiny tuvieron aciertos semejantes, 112 y 91 respectivamente. De cualquier forma, las observaciones humanas y por computadora no son muy dispares salvo en algunos casos puntuales, como se observa en la Ilustración 3, con lo que no es concluyente esta medida para seleccionar el mejor algoritmo.

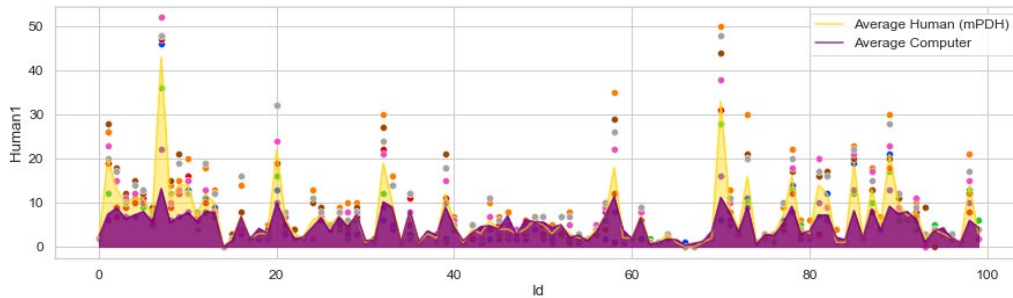


Ilustración 3. Comparación del promedio de personas detectadas por humanos vs computadora para las primeras 100 imágenes

Otro análisis que se realizó fue el nivel de confianza de los algoritmos. La Ilustración 4 demuestra que AWS tiene el mejor nivel de confianza en la detección de personas (Más de 3000 objetos con confianza cercana al 1), con un comportamiento similar de Yolo3; HOGCV es el que peor nivel de confianza presenta, rondando entre 0.1 y 0.2.



Ilustración 4. Distribución de confianza de los algoritmos

Calculando el F1 Score de los cinco algoritmos, se puede tener una idea más completa de la exactitud del algoritmo. La Ilustración 5 muestra los resultados de cada algoritmo de forma gráfica y se pueden apreciar los rectángulos verdes de las personas detectadas con su punto rojo en el centro. Para esta imagen, el F1 Score para cada algoritmo se resume en la Tabla 1, siendo AWS el más alto.

Tabla 1. Resultados de F1 Score para diferentes algoritmos

Alternativa	F1 Score
Yolo3	0.9091
Yolo3 Tiny	0.2857
AWS	1
Azure	0.2857
HOGCV	0.25

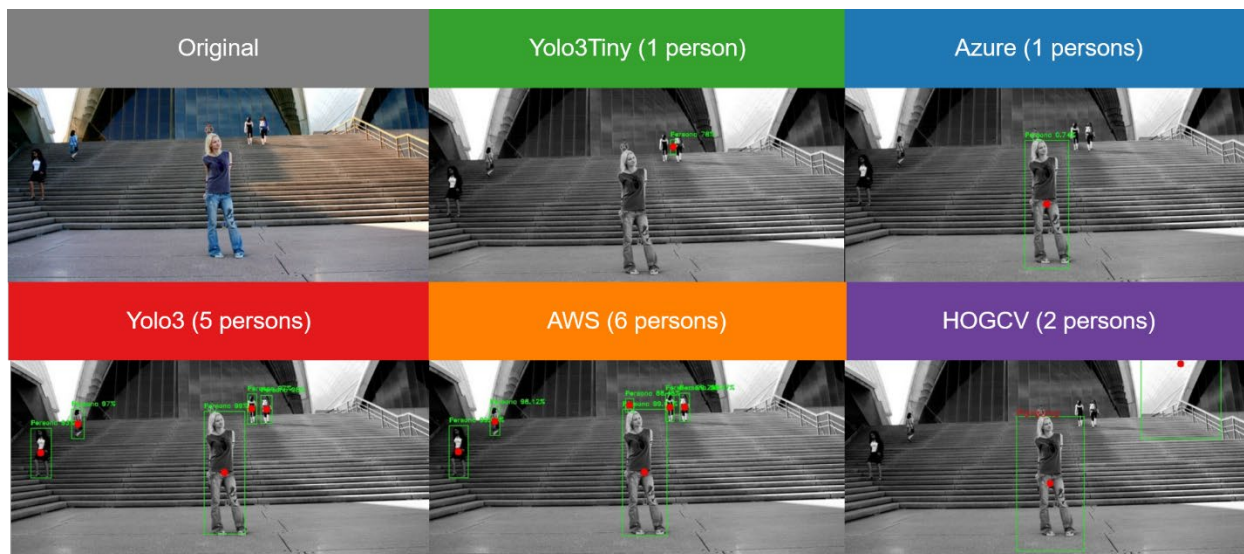


Ilustración 5. Ejemplo de detección de los algoritmos y sus resultados en imágenes

Como sugiere la Tabla 1, el mejor sería AWS, pero en otro ejemplo existe un triple empate (Ilustración 6) donde tres algoritmos obtienen un F1 Score de 1, con lo que no se podría determinar que opción es la mejor.

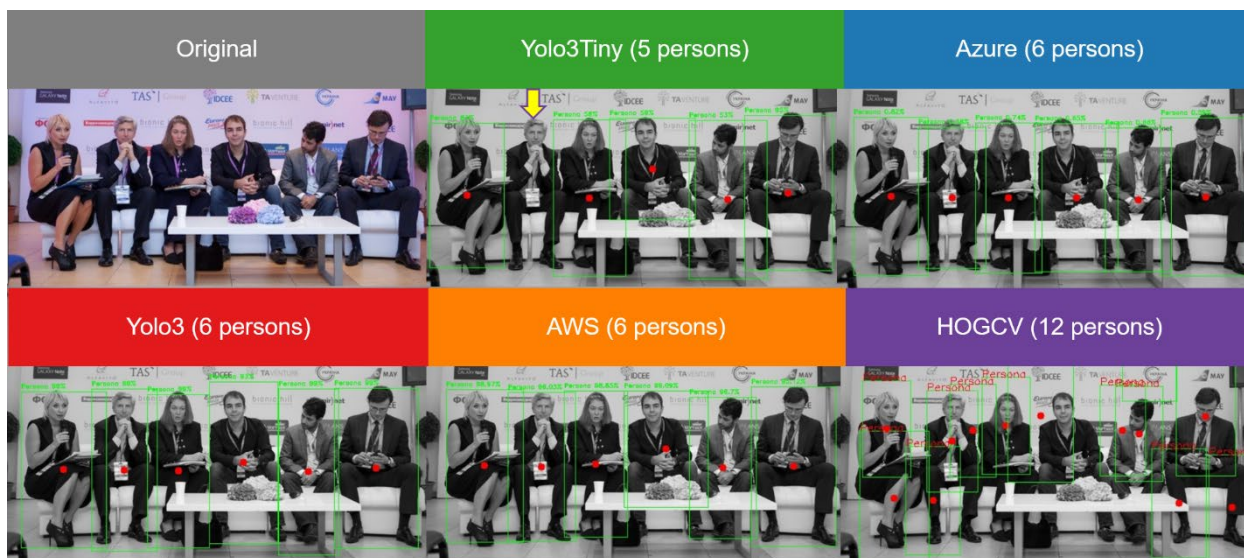


Ilustración 6. Ejemplo donde se presenta un triple empate de F1 Score

Conclusiones

La metodología presentada permite hacer un análisis desde distintas aristas para tener una idea razonada sobre que algoritmo elegir, pudiéndose seguir que el algoritmo con mejor velocidad de procesamiento junto con su mejor F1 Score sería el ideal para seleccionarlo como mejor opción.

También se observó que el F1 Score puede llegar a empatar diversos algoritmos, por lo que no es una medida concluyente pero que puede agrupar los mejores algoritmos para cierta situación.

De las 755 fotos analizadas, se encontró que al menos el algoritmo AWS Rekognition tuvo un alto porcentaje de probabilidad en sus detecciones, sin que sea determinante para concluir que es el mejor de esos cinco.

Por último, el uso de visión por computadora experimentado con al menos cinco alternativas sugiere que es factible para usarse en la detección del número de personas que estén en una parada de camión en un momento determinado. Este número de personas detectadas puede ser utilizado para calcular la demanda en un tiempo t , y con el paso del tiempo, se podría establecer una serie de tiempo para hacer sobre ellos el pronóstico para un tiempo $t+1$, con lo que se podría anticipar la demanda en tiempo casi real.

Las investigaciones futuras que derivan de este trabajo son buscar un índice que permita diferenciar entre alternativas que tienen el mismo F1 Score, y de esa forma seleccionar el mejor algoritmo de forma científica. También, como paso futuro se buscará analizar video en tiempo real y almacenar los datos obtenidos por medio de la visión por computadora, fusionando IIoT con Big Data e inteligencia artificial para la industria de la transportación. Esto no limita que la investigación y el modelo propuesto pueda ser aplicado a una industria diferente.

Referencias

Amazon Web Services. (2020, 08 01). *AWS Rekognition*. Retrieved from AWS: <https://aws.amazon.com/rekognition>

Azure Computer Vision. (2020, 08 1). *Computer vision*. Retrieved from Azure: <https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/>

Contreras-Masse, R., Ochoa-Zezzatti, A., García, V., & Mejía, J. (2019). Application of IoT with haptics interface in the smart manufacturing industry. *International Journal of Combinatorial Optimization Problems and Informatics*, 55-70.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, 1, pp. 886–893.

- Gao, X. (2020). Performance evaluation of automatic object detection with post-processing schemes under enhanced measures in wide-area aerial imagery. *Multimedia Tools and Applications*, 79, 30357–30386.
- Gobierno de la Ciudad de México. (2020, June 1). Retrieved from Red de Transporte Publico: <https://www.rtp.cdmx.gob.mx/>
- Kabakus, A. T. (2019). An Experimental Performance Comparison of Widely Used Face Detection Tools. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, 8, 5–12.
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. *Proceedings of the IEEE international conference on computer vision*, (pp. 2980–2988).
- Mao, Q.-C., Sun, H.-M., Liu, Y.-B., & Jia, R.-S. (2019). Mini-YOLOv3: real-time object detector for embedded applications. *IEEE Access*, 7, 133529–133538.
- Murthy, C. B., Hashmi, M. F., Bokde, N. D., & Geem, Z. W. (2020). Investigations of Object Detection in Images/Videos Using Various Deep Learning Techniques and Embedded Platforms—A Comprehensive Review. *Applied Sciences*, 10, 3280.
- Nazir, M., Jan, Z., & Sajjad, M. (2018). Facial expression recognition using histogram of oriented gradients based transformed features. *Cluster Computing*, 21, 539–548.
- Padilla, R., Netto, S. L., & da Silva, E. A. (2020). A survey on performance metrics for object-detection algorithms. *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, (pp. 237–242).
- Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 7263–7271).
- Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 779–788).
- Translink. (2020, March 21). *Translink Vancouver Bus System*. Retrieved from <https://www.translink.ca/>
- Wang, X., Zheng, S., Yang, R., Luo, B., & Tang, J. (2019). Pedestrian attribute recognition: A survey. *arXiv preprint arXiv:1901.07474*.

Xiao, D., Shan, F., Li, Z., Le, B. T., Liu, X., & Li, X. (2019). A target detection model based on improved tiny-yolov3 under the environment of mining truck. *IEEE Access*, 7, 123757–123764.