

Pontificia Universidad Católica del Perú

Facultad de Ciencias e Ingeniería

**DISEÑO E IMPLEMENTACIÓN DE UNA PLATAFORMA
WEB Y APLICACIÓN MÓVIL CON PROCESAMIENTO
DE IMÁGENES CON INTELIGENCIA ARTIFICIAL
MULTIMODAL EN LA NUBE PARA LA
IDENTIFICACIÓN DE PAGOS, ESTIMACIÓN DE
DIMENSIONES Y OPTIMIZACIÓN LOGÍSTICA EN
SERVICIOS DE DELIVERY URBANO EN LIMA**

Tesis para obtener el título profesional de
Ingeniero de las Telecomunicaciones

Que presenta:

Juan Alfonso Chapoñan Espinoza

Asesor:

Oscar Antonio Díaz Barriga

Lima, junio, 2025

Informe de Similitud

[Por completar]

Índice general

Informe de Similitud	I
1. Presentación del problema de ingeniería	2
1.1. Identificación temática y motivación personal	2
1.1.1. Área de especialización en telecomunicaciones	2
1.1.2. Relación con los estudios realizados	3
1.1.3. Motivación personal y experiencia profesional	4
1.2. Descripción y características del problema	4
1.2.1. Contexto del problema en la industria de <i>delivery</i>	4
1.2.2. Desafío técnico desde la perspectiva de telecomunicaciones . . .	5
1.2.3. Características específicas del problema	5
1.3. Importancia del problema y su solución	5
1.3.1. Perspectiva técnica	5
1.3.2. Perspectiva económica y financiera	6
1.3.3. Perspectiva social y cultural	6
1.3.4. Perspectiva ambiental y de sostenibilidad	6
1.3.5. Perspectiva legal y reglamentaria	6
1.3.6. Perspectiva ética	6
1.4. Impactos previstos y beneficiarios	7
1.4.1. Impactos operacionales	7
1.4.2. Impactos tecnológicos	7
1.4.3. Impactos económicos	7
1.4.4. Beneficiarios directos	7
1.4.5. Beneficiarios indirectos	8

2. Estado del arte o de la cuestión, alternativas de solución al problema o desafío a resolver	9
2.1. Antecedentes de solución semejantes o similares al desafío de ingeniería	9
2.1.1. Sistemas de análisis visual con inteligencia artificial multimodal	9
2.1.2. Aplicaciones de LLMs multimodal en logística y comercio electrónico	12
2.2. Características de soluciones semejantes o similares	13
2.2.1. Arquitecturas de procesamiento predominantes	13
2.2.2. Capacidades de interpretación dimensional	13
2.2.3. Precisión y confiabilidad según implementación	13
2.3. Compendio de tecnologías, herramientas, métodos, modelos utilizados con éxito	14
2.3.1. Modelos de lenguaje multimodal predominantes	14
2.3.2. Frameworks de visión por computadora académicos	14
2.3.3. Tecnologías de infraestructura en la nube	15
2.3.4. Herramientas de desarrollo e integración	15
2.3.5. Metodologías de prompt engineering y optimización	16
2.4. Conjunto de características y especificaciones para la solución óptima .	16
2.4.1. Arquitectura tecnológica óptima	16
2.4.2. Precisión y confiabilidad requeridas	16
2.4.3. Limitaciones y alcances reconocidos	17
3. Estado del arte o de la cuestión, alternativas de solución al problema o desafío a resolver	18
Conclusiones	19
Referencias Bibliográficas	21

[Por completar]

[Por completar]

Capítulo 1

Presentación del problema de ingeniería

La transformación digital en el sector logístico demanda soluciones innovadoras que aprovechen tecnologías emergentes para resolver desafíos operacionales críticos. Esta investigación aborda la problemática de la inexactitud en la determinación de dimensiones de paquetes en la industria de *delivery*, factor que genera ineficiencias significativas en la planificación de rutas y utilización de capacidad vehicular. Mediante la convergencia de aplicaciones IoT, procesamiento de imágenes con inteligencia artificial y arquitecturas distribuidas en la nube, se propone una solución integral que mejora la precisión operacional y democratiza el acceso a tecnologías sofisticadas para empresas de diferentes escalas.

1.1 Identificación temática y motivación personal

1.1.1 Área de especialización en telecomunicaciones

Aplicaciones IoT (*Internet of Things*)

Las aplicaciones IoT constituyen un ecosistema tecnológico integral que integra inteligencia artificial, redes de comunicación y automatización para crear una infraestructura de conectividad ubicua entre objetos físicos y sistemas digitales [1]. Se define como la implementación de una arquitectura de cinco capas interconectadas:

Tabla 1.1: Arquitectura IoT [2].

Capa	Descripción
Percepción	Sensores y dispositivos de captura
Red	Protocolos de comunicación y transmisión
Middleware	Procesamiento y almacenamiento de datos
Aplicación	Servicios e interfaces de usuario
Negocio	Análisis y toma de decisiones

Servicios de Telecomunicaciones para Logística

Los servicios de telecomunicaciones para logística se definen como el conjunto de tecnologías y protocolos de comunicación que operan principalmente en la capa de red del ecosistema IoT, actuando como puente crítico entre la percepción de datos y su procesamiento funcional. Estos servicios garantizan la transmisión eficiente y segura de información tanto estática como móvil durante todas las fases del proceso logístico [2].

Convergencia Tecnológica

La tesis desarrollada en esta investigación representa la convergencia de múltiples disciplinas dentro de la ingeniería de telecomunicaciones:

- **Arquitecturas Distribuidas en la Nube:** Para el procesamiento remoto y almacenamiento escalable
- **Inteligencia Artificial:** Específicamente visión por computadora para el procesamiento automatizado de imágenes

Esta integración tecnológica permite abordar desafíos reales del sector logístico mediante soluciones que aprovechan las capacidades de procesamiento remoto, almacenamiento distribuido y comunicaciones continuas, características fundamentales de los sistemas modernos de telecomunicaciones en el contexto de la transformación digital.

1.1.2 Relación con los estudios realizados

La presente investigación se fundamenta en una progresión curricular especializada que abarca desde los fundamentos del desarrollo web hasta la implementación de soluciones IoT avanzadas, estableciendo una relación directa y sistemática con tres cursos clave que proporcionan las competencias técnicas necesarias.

TEL131 Ingeniería Web para Telecomunicaciones

Este curso proporciona la base tecnológica para desarrollar la capa de aplicación e interfaces de gestión en soluciones de logística inteligente. Se abordan fundamentos de programación, desarrollo web con conexión a bases de datos, y modelado relacional con SQL, aplicables en *dashboards* para monitoreo en tiempo real, interfaces de control y manejo de datos IoT.

TEL137 Gestión de Servicios de TICs

Este curso se enfoca en la gestión de servicios dentro del ecosistema IoT, brindando competencias para desarrollar infraestructuras seguras, escalables y robustas. A través de *frameworks* modernos y servicios web, se construyen sistemas capaces de optimizar rutas y asignar recursos inteligentemente.

1TEL05 Servicios y Aplicaciones para IoT

Este curso se centra en la construcción de la capa de aplicación IoT y su integración con la nube, facilitando el desarrollo de soluciones logísticas orientadas al usuario final. Se abordan competencias en aplicaciones móviles conectadas a servicios SaaS, útiles para rastreo de productos y alertas inteligentes.

1.1.3 Motivación personal y experiencia profesional

La motivación personal para abordar esta problemática combina una vocación por la automatización de procesos con un interés social en mejorar la eficiencia empresarial. La experiencia profesional durante la carrera, desde almacenero hasta asistente logístico, brindó una comprensión directa de los desafíos operacionales, destacando la importancia del cumplimiento de tiempos en cada etapa del proceso logístico.

1.2 Descripción y características del problema

1.2.1 Contexto del problema en la industria de *delivery*

La industria de servicios de *delivery* y logística de última milla ha experimentado un crecimiento exponencial en los últimos años, impulsada por el auge del comercio electrónico y los cambios en los hábitos de consumo de la población [3].

Tabla 1.2: Proyecciones del Mercado de delivery en Perú.

Año	Crecimiento (%)	Ingresos (millones USD)
2024	—	2,500
2029	11.03	2,951

Uno de los principales desafíos que enfrentan las empresas de *delivery* es obtener información precisa sobre las dimensiones y características de los paquetes que deben recoger y entregar.

1.2.2 Desafío técnico desde la perspectiva de telecomunicaciones

Desde la perspectiva de la ingeniería de telecomunicaciones, el problema central radica en la necesidad de desarrollar un sistema distribuido que permita el procesamiento automatizado de imágenes capturadas por dispositivos móviles para la determinación precisa de dimensiones de paquetes.

1.2.3 Características específicas del problema

El problema presenta características específicas que lo hacen particularmente complejo desde el punto de vista técnico y operacional:

- Gran variabilidad en dimensiones, formas y características físicas de los paquetes
- Limitación de capacidad de los motorizados
- Necesidad de confiabilidad en los procesos de verificación
- Requisito de escalabilidad del sistema

1.3 Importancia del problema y su solución

1.3.1 Perspectiva técnica

Resolver este problema es clave por el uso de tecnologías emergentes que transforman telecomunicaciones y computación distribuida. El procesamiento en la nube ha madurado para realizar análisis complejos de imágenes sin necesidad de ejecución local en dispositivos [4], [5].

1.3.2 Perspectiva económica y financiera

La automatización en medición de paquetes mejora rentabilidad y competitividad logística, generando ahorros significativos y optimizando recursos de transporte [6].

Figura 1.1: Gasto en tecnologías de la información - América Latina.

1.3.3 Perspectiva social y cultural

Socialmente, la solución mejora la calidad de vida de trabajadores logísticos y usuarios finales. La automatización de tareas repetitivas permite enfocar recursos humanos en actividades de mayor valor, mejorando satisfacción laboral y desarrollo profesional.

1.3.4 Perspectiva ambiental y de sostenibilidad

Abordar este problema es clave para reducir emisiones de gases de efecto invernadero y usar recursos energéticos eficientemente. La optimización de rutas, basada en medidas precisas de paquetes, permite planificar trayectos más cortos, disminuyendo consumo de combustible y emisiones de CO₂.

1.3.5 Perspectiva legal y reglamentaria

En Perú, la medición automatizada de paquetes con IoT debe cumplir la Ley N.º 29733, que exige consentimiento previo, finalidad clara, calidad y seguridad en el tratamiento de datos [7].

1.3.6 Perspectiva ética

La automatización plantea dilemas éticos sobre responsabilidad social, equidad digital, sostenibilidad ambiental, privacidad de datos y rendición de cuentas. Esta visión ética integral asegura que la tecnología respete valores humanos y apoye un desarrollo sostenible.

1.4 Impactos previstos y beneficiarios

1.4.1 Impactos operacionales

La implementación tendrá impactos operacionales significativos, mejorando la eficiencia mediante tecnologías IoT, IA y visión artificial [8], [9].

Reducción de tiempos de procesamiento

La medición automatizada elimina procesos manuales, reduciendo el tiempo requerido para registrar, verificar y procesar solicitudes [4].

Optimización de rutas de entrega

Con datos precisos sobre dimensiones, los sistemas pueden planificar rutas más eficientes considerando capacidad de carga, tiempo y distancia.

1.4.2 Impactos tecnológicos

La implementación tendrá un impacto tecnológico significativo en telecomunicaciones e IoT, estableciendo nuevos paradigmas en procesamiento de imágenes para logística.

1.4.3 Impactos económicos

Los impactos económicos se reflejan desde la productividad individual hasta la competitividad sectorial, aprovechando procesamiento de imágenes con IA en la nube para optimizar entregas urbanas.

1.4.4 Beneficiarios directos

Empresas de *delivery* y logística

Son los principales beneficiarios, mejorando eficiencia y rentabilidad gracias al procesamiento de imágenes con IA en la nube.

Motorizados y personal operativo

Mejoran productividad y condiciones laborales mediante rutas optimizadas que permiten más entregas en menos tiempo.

Clientes emisores de paquetes

Experimentan mayor fiabilidad y transparencia con rutas inteligentes que optimizan entregas y recojo.

Destinatarios de entregas

Reciben un servicio más rápido y predecible, cumpliendo estándares de *quick commerce* en menos de 90 minutos.

1.4.5 Beneficiarios indirectos**Sector de Telecomunicaciones**

Este sector verá un aumento en la demanda de servicios especializados y mejoras en infraestructura.

Industria de Desarrollo de Software

Se generarán nuevas oportunidades y mayor demanda de talento especializado en aplicaciones móviles, NoSQL y microservicios.

Medio Ambiente y Sociedad

La optimización de rutas reducirá emisiones de CO₂ y consumo de combustible, contribuyendo a un entorno urbano más saludable.

Ecosistema de Innovación Tecnológica

La solución fortalecerá la innovación al demostrar cómo tecnologías emergentes resuelven problemas reales.

Capítulo 2

Estado del arte o de la cuestión, alternativas de solución al problema o desafío a resolver

2.1 Antecedentes de solución semejantes o similares al desafío de ingeniería

2.1.1 Sistemas de análisis visual con inteligencia artificial multimodal

Caso 1: GPT-4 Vision para análisis de objetos (OpenAI, Estados Unidos, 2023)

GPT-4 Vision, lanzado en septiembre de 2023, representa el primer modelo de lenguaje de gran escala con capacidades multimodal nativas de OpenAI. El sistema demuestra capacidades avanzadas de interpretación visual con un 95 % de precisión en reconocimiento de objetos comunes y 78.5 % de efectividad en análisis de gráficos complejos según evaluaciones técnicas independientes [10].

En términos de estimación dimensional, GPT-4V utiliza razonamiento contextual para comparar tamaños relativos entre objetos, empleando elementos conocidos como referencias de escala. Las evaluaciones técnicas reportan una precisión de $\pm 15\text{--}25\%$ de error en estimaciones dimensionales cuando se proporcionan referencias visuales adecuadas, mejorando a $\pm 10\text{--}20\%$ bajo condiciones de iluminación controlada.

Figura 2.1: Consulta a GPT-4 para análisis de múltiples imágenes.

La arquitectura se basa en un *transformer* multimodal que integra un codificador visual

especializado con el modelo de lenguaje GPT-4, estimado en 1.76 trillones de parámetros. El sistema procesa imágenes de hasta 2048×2048 píxeles en formatos JPEG, PNG, GIF y WebP.

Las evaluaciones técnicas del sistema revelan un rendimiento variable según el contexto de aplicación:

- *MMMU Benchmark*: 56.8 % en tareas multimodal complejas
- *MathVista*: 49.9 % en razonamiento visual-matemático
- *AI2D*: 78.2 % en interpretación de diagramas técnicos
- *ChartQA*: 78.5 % en análisis de gráficos y visualizaciones

Las limitaciones incluyen:

- Sin calibración externa: Incremento del error a $\pm 30\text{--}50\%$
- Sensibilidad ambiental: Degradación con variaciones en iluminación
- Limitaciones de escala: Rendimiento reducido en objetos menores a 5cm
- Objetos problemáticos: Dificultades con elementos transparentes

Caso 2: Claude 3 Vision

Claude 3, lanzado en marzo de 2024 en sus variantes Haiku, Sonnet y Opus, establece un nuevo estándar en interpretación multimodal con un enfoque específico en razonamiento avanzado sobre contenido visual complejo. El sistema demuestra capacidades superiores a GPT-4V en interpretación de gráficos y documentos técnicos, alcanzando 86.8 % en el *benchmark* MMLU y 60.1 % en problemas matemáticos complejos.

Figura 2.2: Identificación de objetos visualmente usando modelos de Claude 3.

La arquitectura del modelo incorpora una ventana de contexto extendida de 200,000 tokens que incluye contenido visual, permitiendo el análisis simultáneo de múltiples imágenes y documentos dentro de una sola conversación.

Figura 2.3: Solicitud de reconocimiento de una imagen y reorganización en formato JSON.

Para estimación dimensional, Claude 3 utiliza razonamiento contextual sofisticado que combina reconocimiento de objetos conocidos con análisis proporcional de elementos en la imagen. El sistema puede interpretar planos técnicos con dimensiones especificadas y extrapolar esta información para estimar las dimensiones de objetos fotografiados, logrando precisiones de $\pm 20\text{--}30\%$ en estimaciones sin referencias calibradas externas.

Las implementaciones documentadas incluyen:

- Análisis de especificaciones técnicas
- Interpretación de diagramas de embalaje
- Auditoría visual de inventarios
- Análisis de documentos comerciales

Ventajas competitivas: ventana de contexto extendida (200K vs 128K tokens), mejor comprensión de documentos complejos, razonamiento espacial más avanzado, y menor tendencia a alucinaciones en datos técnicos críticos.

Caso 3: Gemini 2.0 Flash

Gemini 2.0 Flash, lanzado en diciembre de 2024, representa la segunda generación de modelos multimodal de Google con optimizaciones específicas para velocidad de procesamiento y análisis visual en tiempo real [11]. El modelo incorpora una arquitectura *transformer* multimodal de segunda generación optimizada para respuestas rápidas, logrando velocidades hasta $2\times$ superiores a Gemini 1.0.

Figura 2.4: Se solicita a Gemini reconocer las imágenes y encontrar una relación entre ellas.

El sistema soporta modalidades múltiples incluyendo texto, imagen, audio y video, con capacidad de procesamiento de imágenes de hasta 30MB en formatos JPEG, PNG, GIF, WebP, PDF, SVG y HEIC. La ventana de contexto se expande masivamente a 2 millones de tokens.

Gemini 2.0 Flash integra Google Lens como módulo nativo, eliminando la arquitectura de integración externa de generaciones anteriores. El sistema demuestra precisión mejorada en estimaciones dimensionales, alcanzando márgenes de $\pm 8\text{--}15\%$ con objetos de referencia en productos estándar, $\pm 5\text{--}12\%$ en condiciones de laboratorio controlado, y $\pm 12\text{--}25\%$ en uso típico de usuarios reales [12].

Implementaciones específicas en logística incluyen:

- Análisis de inventario en tiempo real para Google Shopping
- Medición automatizada de productos para Google Merchant Center
- Optimización de embalaje en centros de distribución
- Análisis de fotografías de productos para Google Lens Shopping

Caso 4: Vision Transformer (ViT) para reconocimiento y dimensionado de productos

Vision Transformer (ViT), introducido por Dosovitskiy et al. en Google Research (2020) y publicado en ICLR 2021, revolucionó el campo de *computer vision* al demostrar que arquitecturas transformer puras pueden superar las redes neuronales convolucionales tradicionales en tareas de reconocimiento de imágenes. El modelo alcanza 94.2% de precisión en *ImageNet classification* cuando se *pre-entrena* en *datasets* suficientemente grandes [13].

La arquitectura divide imágenes en *patches* de 16×16 píxeles que se procesan como secuencias, similar al procesamiento de tokens en modelos de lenguaje. Esta metodología permite *transfer learning* eficiente para nuevas categorías de productos.

Las implementaciones académicas posteriores han demostrado aplicabilidad directa en *retail analytics*. DINOv2 (Meta AI Research, 2023) introduce *self-supervised learning* que elimina la dependencia de datasets etiquetados masivos, logrando representaciones visuales robustas especialmente efectivas para reconocimiento de productos [14].

Para estimación dimensional, las investigaciones académicas reportan precisiones de $\pm 12\text{--}18\%$ en objetos con referencias visuales, utilizando visual *reasoning* sobre *features transformer* que capturan relaciones espaciales complejas.

2.1.2 Aplicaciones de LLMs multimodal en logística y comercio electrónico

Florence (Microsoft Research, 2021) establece un framework multimodal que combina *vision transformers* con capacidades de lenguaje natural, demostrando aplicabilidad directa para tareas que requieren comprensión semántica de productos y estimación de propiedades físicas.

Las evaluaciones académicas independientes confirman escalabilidad para procesamiento de millones de productos, con implementaciones que mantienen eficiencia computacional mediante técnicas de *patch embedding* optimizadas.

2.2 Características de soluciones semejantes o similares

2.2.1 Arquitecturas de procesamiento predominantes

Las soluciones analizadas convergen en arquitecturas de procesamiento que combinan múltiples enfoques tecnológicos. Los sistemas comerciales líderes utilizan *transformers* multimodal que integran vision encoders especializados (CLIP, ALIGN, ViT) con *large language models* para interpretación semántica avanzada.

Las implementaciones actuales emplean *attention mechanisms cross-modal* que permiten correlación directa entre características visuales y comprensión textual, facilitando estimación dimensional mediante razonamiento contextual [13].

2.2.2 Capacidades de interpretación dimensional

Las capacidades de interpretación dimensional se fundamentan en:

1. **Reconocimiento automático de referencias de escala:** Los modelos identifican objetos conocidos (monedas, tarjetas, manos humanas) para establecer proporciones relativas.
2. **Análisis de perspectiva y profundidad visual:** Utiliza *depth estimation* implícito derivado de características de textura y sombras.
3. **Razonamiento contextual sofisticado:** Combina reconocimiento de objetos conocidos con análisis proporcional de elementos en la imagen [14].

2.2.3 Precisión y confiabilidad según implementación

La precisión varía significativamente según la implementación y condiciones operacionales:

Tabla 2.1: Comparativa de precisión por tipo de sistema.

Tipo de Sistema	Condiciones Óptimas	Condiciones Reales
LLMs multimodal	$\pm 10\text{--}25\%$	$\pm 15\text{--}30\%$
Vision Transformers	$\pm 12\text{--}18\%$	$\pm 15\text{--}25\%$
Sistemas con hardware	$\pm 2\text{--}5\text{mm}$	$\pm 5\text{--}10\text{mm}$

La degradación de precisión con iluminación deficiente, ángulos subóptimos y *back-grounds* complejos representa un desafío común.

2.3 Compendio de tecnologías, herramientas, métodos, modelos utilizados con éxito

2.3.1 Modelos de lenguaje multimodal predominantes

Modelos de gran escala comerciales:

- GPT-4 Vision (OpenAI) - 1.76 trillones de parámetros
- Claude Sonnet 4 (Anthropic) - Optimizado para razonamiento visual
- Gemini 2.0 Flash (Google) - Optimizaciones de velocidad
- LLaVA - Alternativas *open-source*

Arquitecturas especializadas predominantes:

- CLIP (*Contrastive Language-Image Pre-training*)
- BLIP-2 (*Bootstrapped vision-language pre-training*)
- InstructBLIP - Seguimiento de instrucciones específicas
- MiniGPT-4 - Eficiencia en recursos limitados

2.3.2 Frameworks de visión por computadora académicos

Vision Transformers (ViT) representan el estado del arte en análisis visual académico:

- ViT-Base: 86 millones de parámetros
- ViT-Large: 307 millones de parámetros
- ViT-Huge: 632 millones de parámetros

Modelos híbridos incluyen DeiT (*Data-efficient image Transformers*), Swin Transformer optimizado para imágenes de alta resolución, y EfficientViT diseñado para implementaciones móviles.

2.3.3 Tecnologías de infraestructura en la nube

Plataformas de IA como servicio:

- OpenAI API con GPT-4 Vision endpoints
- Anthropic Claude API para análisis multimodal
- Google Vertex AI con modelos Gemini integrados
- Azure Cognitive Services
- AWS Bedrock

Servicios de computación distribuida:

- AWS Lambda - Procesamiento *serverless*
- Google Cloud Run - Containerización
- Azure Container Instances - Escalabilidad elástica
- Kubernetes - Orquestación de microservicios

2.3.4 Herramientas de desarrollo e integración

Bibliotecas de integración:

- LangChain - Framework para aplicaciones LLM
- LlamaIndex - *Retrieval-Augmented Generation*
- Haystack - Pipelines *end-to-end*
- Transformers (HuggingFace) - Acceso a modelos pre-entrenados

SDKs para desarrollo móvil:

- React Native con plugins para APIs de IA
- Flutter con packages para computer vision
- Swift/Kotlin con SDKs nativos
- Progressive Web Apps para acceso universal

2.3.5 Metodologías de prompt engineering y optimización

Técnicas avanzadas de diseño de prompts:

- *Few-shot learning* para medición dimensional
- *Chain-of-thought prompting* para razonamiento paso a paso
- *Structured output formatting* para consistencia
- *Multi-turn conversation* para refinación iterativa

Estrategias de optimización:

- *Fine-tuning* con datos del dominio logístico
- *Retrieval-Augmented Generation* (RAG)
- *In-context learning* con ejemplos calibrados
- LoRA y Adapters para personalización eficiente

2.4 Conjunto de características y especificaciones para la solución óptima

2.4.1 Arquitectura tecnológica óptima

Basándose en el análisis de soluciones existentes, la arquitectura óptima debe combinar:

- Modelos de lenguaje multimodal de última generación
- Infraestructura *cloud* escalable
- Procesamiento distribuido eficiente
- Interfaces móviles intuitivas

2.4.2 Precisión y confiabilidad requeridas

Para aplicaciones logísticas comerciales, el sistema debe alcanzar:

- Precisión de $\pm 15\text{--}20\%$ en condiciones reales de uso

- Mejora a $\pm 10\text{--}15\%$ con referencias de escala adecuadas
- Tiempos de respuesta inferiores a 5 segundos
- Disponibilidad superior al 99.5 %

Esta precisión es suficiente para categorización de tarifas de envío, optimización básica de carga vehicular, y estimaciones de costos logísticos sin requerir inversión en hardware especializado.

2.4.3 Limitaciones y alcances reconocidos

El sistema está diseñado para:

- Estimaciones dimensionales aproximadas para categorización logística
- Paquetes regulares de e-commerce (5cm a 2 metros)
- Condiciones de iluminación mínima adecuada
- Mercado peruano inicialmente, con escalabilidad regional

Limitaciones reconocidas:

- No apto para aplicaciones que requieran precisión milimétrica
- Degradación con objetos transparentes o altamente reflectivos
- Requiere iluminación mínima adecuada
- Formas extremadamente irregulares presentan mayor error

Capítulo 3

Estado del arte o de la cuestión, alternativas de solución al problema o desafío a resolver

Conclusiones

[Por completar]

Referencias Bibliográficas

- [1] «IoT-based TPL whole supply chain logistics information system model,» *Proceedings - International Conference on Machine Learning and Cybernetics*, vol. 4, págs. 1758-1762, jul. de 2013, ISSN: 21601348. DOI: 10.1109/ICMLC.2013.6890882. dirección: <https://ieeexplore.ieee.org/document/6890882>.
- [2] «Applications of Computer Internet of Things Technologies in Logistics,» *2nd IEEE International Conference on Advanced Technologies in Intelligent Control, Environment, Computing and Communication Engineering, ICATIECE 2022*, 2022. DOI: 10.1109/ICATIECE56365.2022.10047270. dirección: <https://ieeexplore.ieee.org/document/10047270>.
- [3] R. TLW, *Desafíos y soluciones en la distribución de última milla: ¿Cómo adaptarse al futuro?* Feb. de 2025. dirección: <https://thelogisticsworld.com/logistica-y-distribucion/desafios-y-soluciones-en-la-distribucion-de-ultima-milla-en-entornos-urbanos/>.
- [4] «Design and actualization of IoT-based intelligent logistics system,» *IEEE International Conference on Industrial Engineering and Engineering Management*, págs. 2245-2248, 2012, ISSN: 2157362X. DOI: 10.1109/IEEM.2012.6838146.
- [5] «Research on Cloud Logistics-based One-stop Service Platform for Logistics Center,» *Proceedings of the 2012 IEEE 16th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2012*, págs. 558-563, 2012. DOI: 10.1109/CSCWD.2012.6221873.
- [6] J. Krysińska, *On-premise vs. cloud: definition & differences* / NordLayer, ago. de 2024. dirección: <https://nordlayer.com/blog/on-premise-vs-cloud-differences/>.
- [7] E. Perú, «REGLAMENTO DE LA LEY N° 29733 LEY DE PROTECCIÓN DE DATOS PERSONALES,» 2973. dirección: <https://cdn.www.gob.pe/uploads/document/file/272360/Ley%20N%C2%BA%2029733.pdf.pdf?v=1618338779>.
- [8] R. TLW, *Supervisión de seguridad en tiempo real en la cadena de suministro: Innovación con IoT*, nov. de 2024. dirección: <https://thelogisticsworld.com/tecnologia/el-uso-de-iot-para-la-supervision-de-seguridad-en-tiempo-real-en-la-cadena-de-suministro/>.

- [9] «Green and Network-Aware Smart IoT Logistics Applications,» *2023 International Conference on Information Technology: Cybersecurity Challenges for Sustainable Cities, ICIT 2023 - Proceeding*, págs. 441-446, 2023. DOI: 10.1109/ICIT58056.2023.10225953.
- [10] «MM-Vet: Evaluating Large Multimodal Models for Integrated Capabilities,» 2024. dirección: <https://huggingface..>
- [11] «Gemini: A Family of Highly Capable Multimodal Models,» 2025.
- [12] G. Team, «Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities.,» 2025.
- [13] «AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE,» dirección: <https://github.com/>.
- [14] «DINOv2: Learning Robust Visual Features without Supervision,» dirección: <https://openreview.net/forum?id=a68SUt6zFt>.