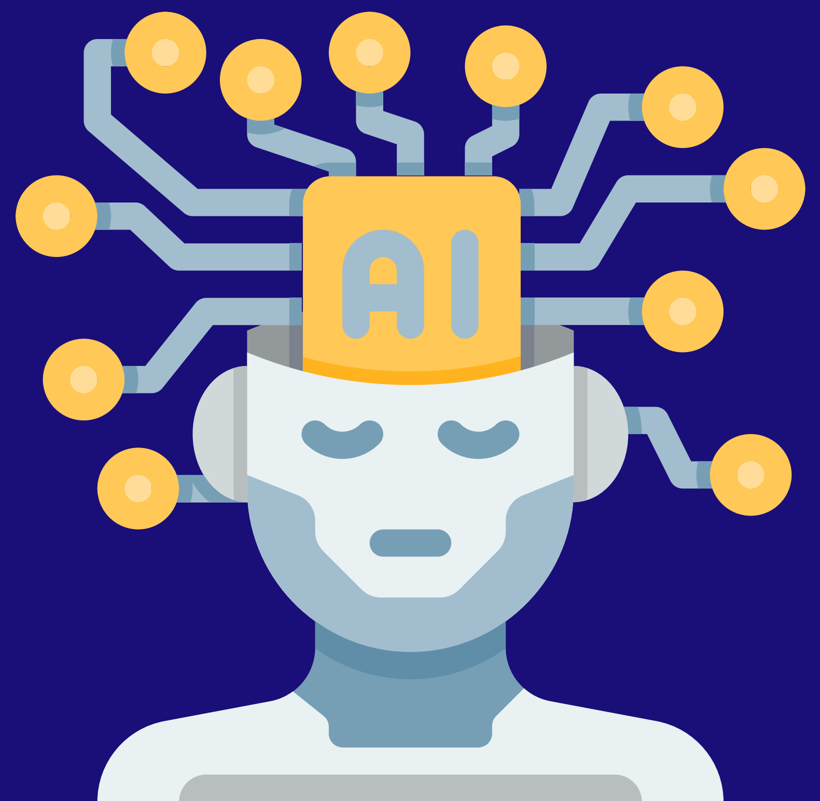
PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

# Multi-modal proxy learning towards personalized visual multiple clustering

Jiawei Yao, Qi Qian, Junhua Hu

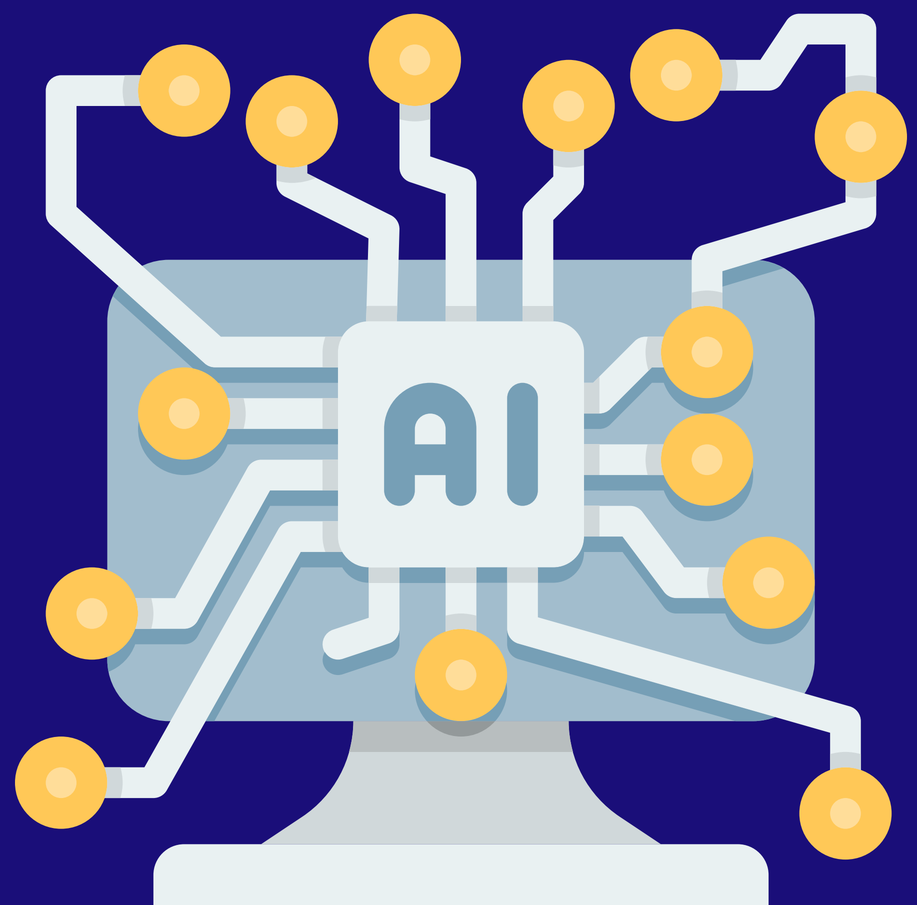



Benjamín Díaz
Ilan San Martín
Sebastían Terrazas

**IIC3666**

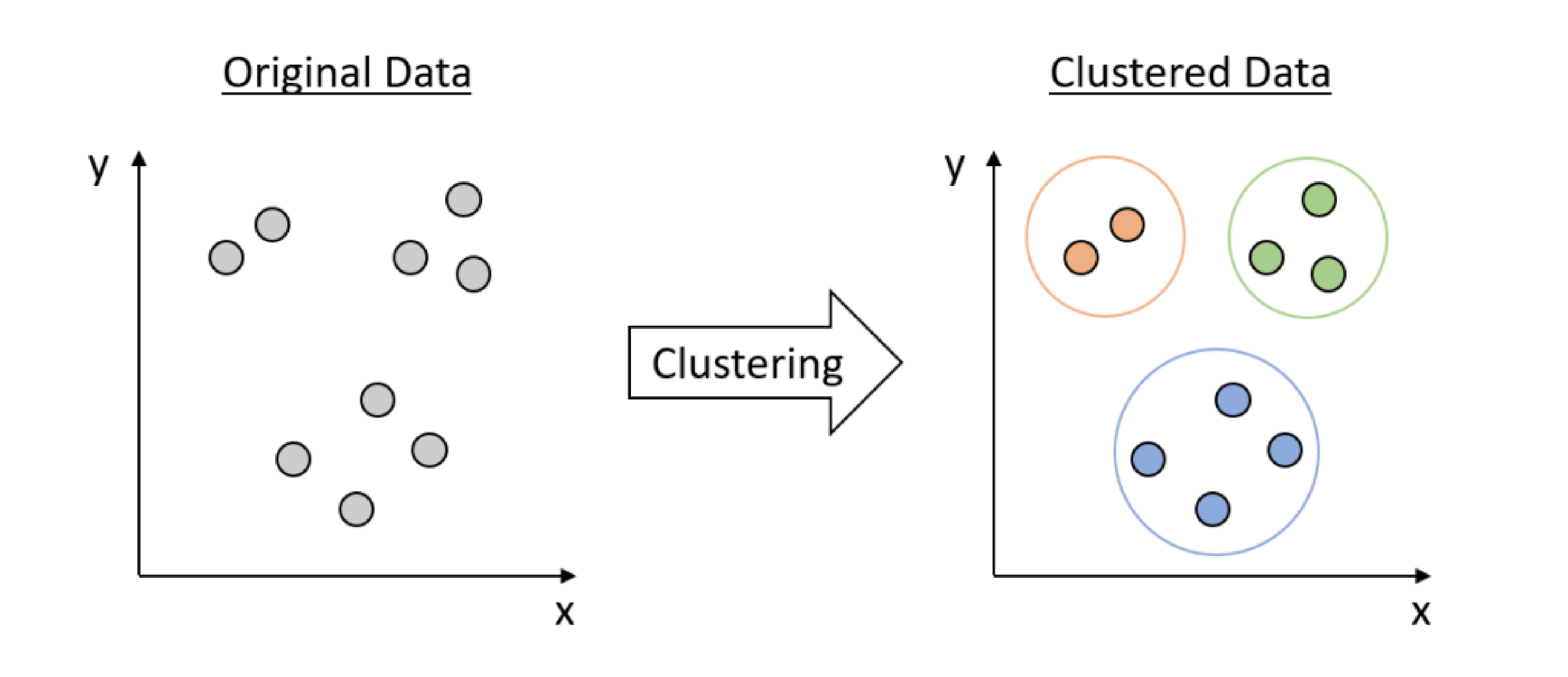
# Índice

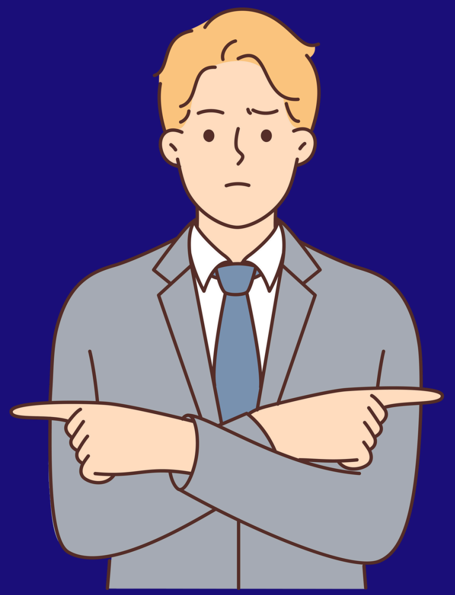




# Contexto

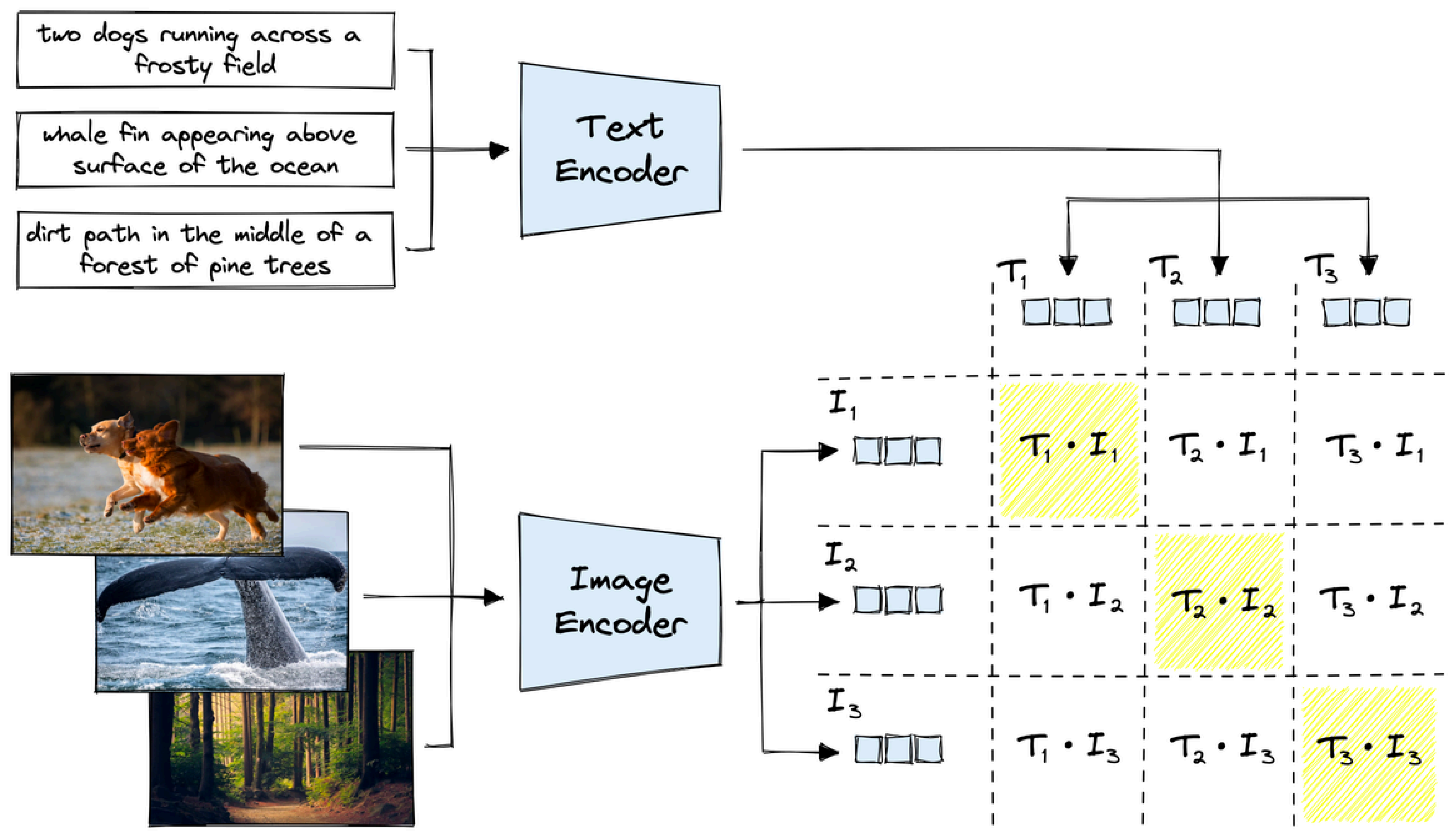
- **Clustering:** Técnica de agrupamiento de datos según similitud
- **Multiple Clustering**: Generar **múltiples agrupaciones** posibles para el mismo conjunto de datos.
- La llegada de **Deep Learning** ayuda en gran medida al **rendimiento** del clustering.
- El **problema**: ¿Cuáles **agrupaciones** son **relevantes**?



# Problema de recomendación

- Enfoque de aprendizaje automático del interés del usuario y agrupación adecuada.
- Multi-Map (Multi-Modal Proxy Learning): Combinación modelos multimodales (CLIP) y modelos de lenguaje (GPT-4).

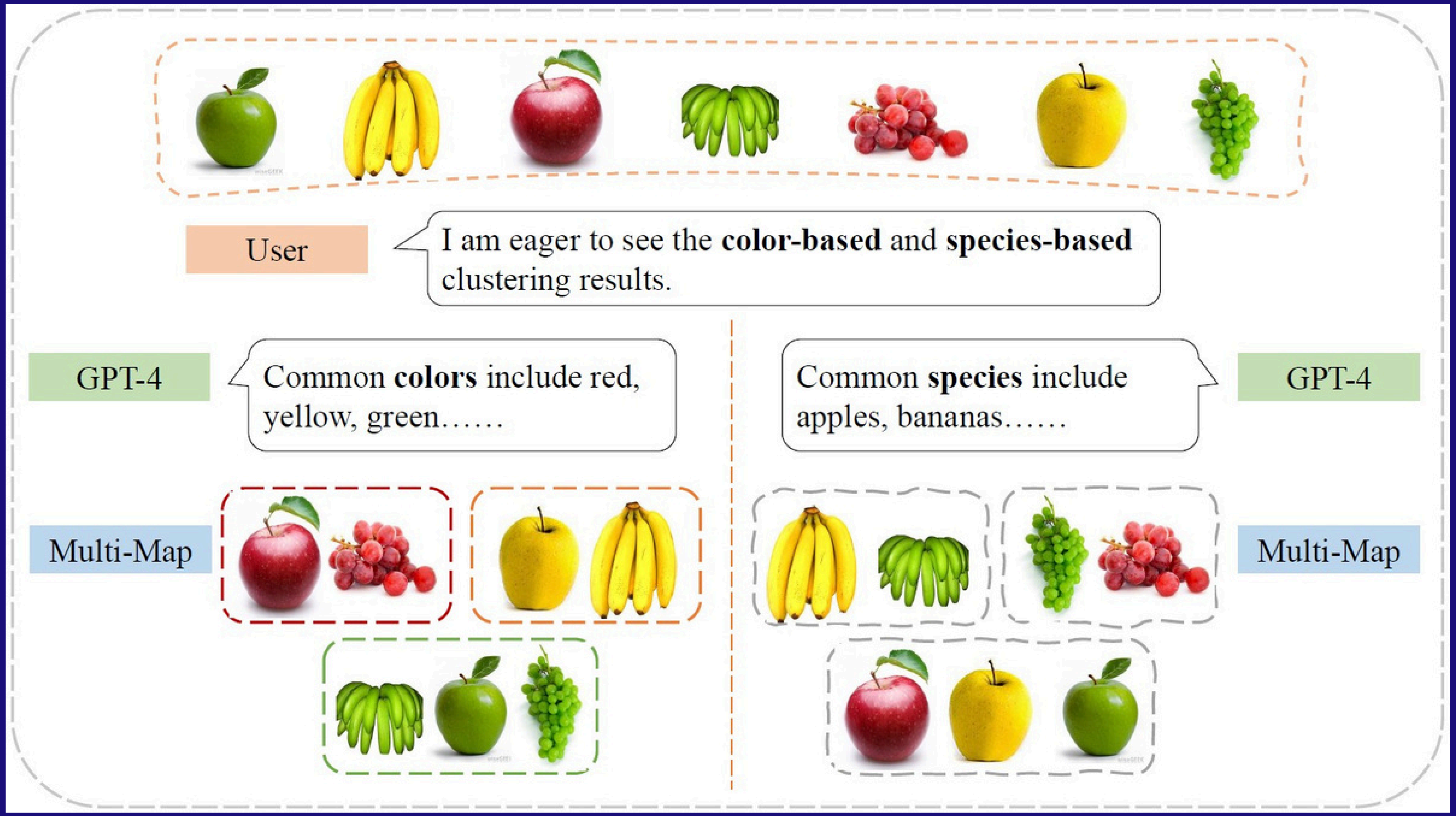



OpenAI

GPT - 4

# Contribución

- Clustering múltiple **Personalizado** y **NO Supervisado**
- Basado en **preferencias del usuario** en **lenguaje natura**l
- **SIN** necesidad de **etiquetas explícitas**.
- **Evita** la **revisión manual** de múltiples clusterings



# Estado del arte y marco teórico


**MSC**

**Algoritmo clásico con hand-crafted features**
Genera múltiples clusterings sin asegurar diversidad

**2017**

**2020**

**ENRC**

**Usa un autoencoder para aprender las features**
Penaliza redundancia en clustering

**CLIP**

**Modelo multimodal**
Backbone visual-textual de Multi MaP, usándolo como encoder

**2021**

**2022**

**CoOp**

**Aprenden prompts óptimos en CLIP**
Inspiración para el aprendizaje de proxies textuales

**AugDMC**

**Aplica data augmentation sobre las imágenes**
Estado del arte de deep multiple clustering

**2023**

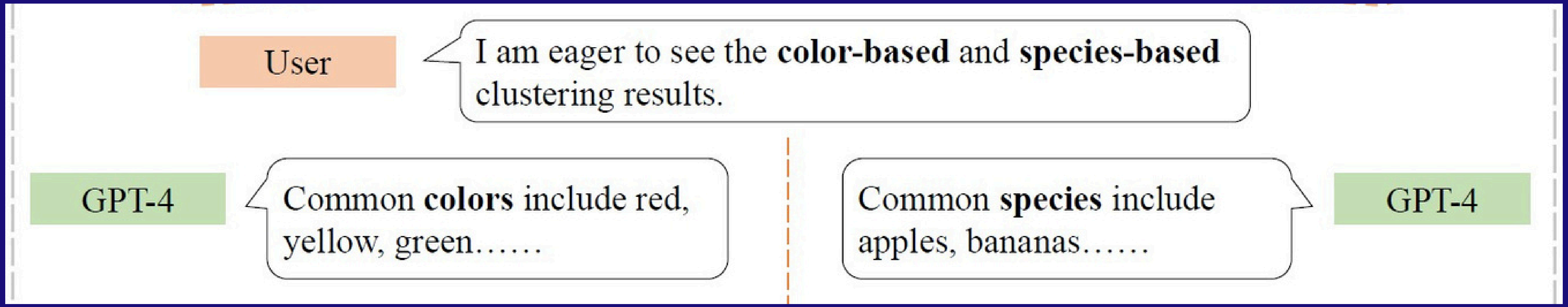
Multiple Clustering

Componentes clave

# Solución

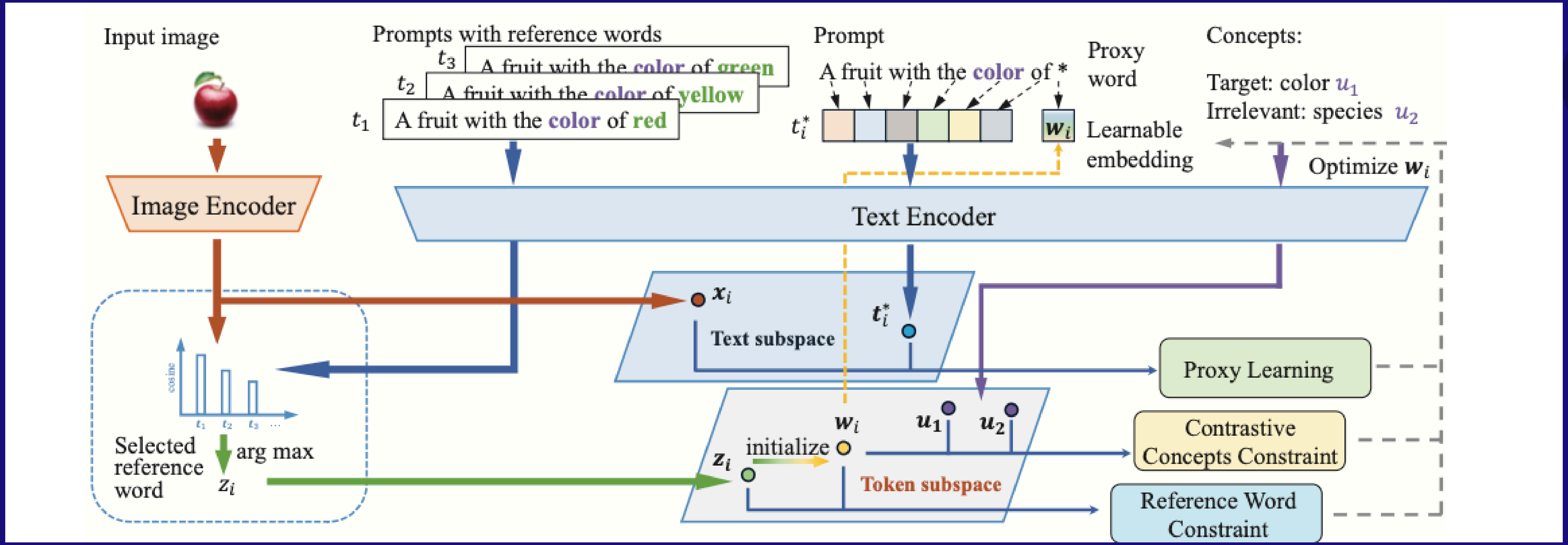
## Flujo de Multi-Map

1. El usuario proporciona un concepto general (ej: "color", "especie").
2. El modelo GPT-4 genera automáticamente palabras concretas relacionadas (ej: "rojo", "verde", etc.).
3. Se codifican esas palabras generando múltiples prompts textuales. Para cada imagen, se selecciona el prompt más adecuado según similitud.



# Solución

## Flujo de Multi-Map




4. Se inicializa un proxy word (vector representativo) con la palabra seleccionada y se optimiza:
    - Para maximizar la similitud con la representación visual de la imagen.
    - Para mantener proximidad al prompt específico y al concepto general proporcionado.
    - Para minimizar confusión entre diferentes conceptos (restricción contrastiva).
5. Finalmente, se ejecuta k-means sobre estos vectores aprendidos, generando agrupaciones personalizadas según interés del usuario.

# Solución

## Pérdidas y Ecuaciones Clave

$$\mathcal{L}\left(\mathbf{w}_i\right) = \mathcal{L}_{\text{align}} + \alpha\mathcal{L}_{\text{reference}} + \beta\mathcal{L}_{\text{concept}}$$

$$\mathcal{L}_{\text{align}} = -\langle f(x_i), h(t_i^*)\rangle$$

$$\mathcal{L}_{\text{reference}} = \|\mathbf{w}_i - \mathbf{z}_i\|_2^2$$

$$\mathcal{L}_{\text{concept}} = -\log \frac{\exp(\mathbf{w}_i^\top \mathbf{u}_w)}{\sum_j \exp(\mathbf{w}_i^\top \mathbf{u}_j)}$$

$x_i$ : Imagen ej: red apple

$t_i^*$ : Prompt para CLIP
ej: "a fruit with the color of red"

$\mathbf{w}_i$ : Proxy Imagen
init como "red", cambia al aprender

$\mathbf{z}_i$ : Palabra referencia
ej: "red"

$\mathbf{u}_w$ : Concepto general
ej: "color"

$\{\mathbf{u}_j\}$ : Otros conceptos
ej. "species", "shape"

# Solución

## Experimentación

- CLIP y GPT-4 no se fine-tunean: sólo **entrenamos** los $\mathbf{w}_i$
- Optimización con **Adam**, momentum = 0.9, durante 1000 epochs.
- Tasa de aprendizaje (**Learning Rate**) seleccionada entre (0.1 a 0.0005).
- Decaimiento de peso (**Weight decay**) seleccionado entre (0.0005 a 0).
- Parámetros de restricciones: **α y β** en rango (0.0 a 1.0). Se escogen por grid-search validando la pérdida total en el conjunto de entrenamiento
- Evaluación usando métricas: Normalized Mutual Information (**NMI**) y Rand Index (**RI**). Miden qué tanta "información" comparten dos particiones; y la proporción de pares de puntos que están de acuerdo en las dos particiones, respectivamente

# Resultados

## Multi-MAP vs Métodos existentes

| Dataset | Clustering | MSC NMI | MSC RI | MCV NMI | MCV RI | ENRC NMI | ENRC RI | iMClusts NMI | iMClusts RI | AugDMC NMI | AugDMC RI | Multi-MaP NMI | Multi-MaP RI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ALOI | Color | 0.1563 | 0.3428 | 0.6982 | 0.7439 | 0.9833 | 0.9892 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | Shape | 0.2968 | 0.5199 | 0.7359 | 0.8261 | 0.9732 | 0.9861 | 0.9963 | 0.9989 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Fruit | Color | 0.6886 | 0.8051 | 0.6266 | 0.7685 | 0.7103 | 0.8511 | 0.7351 | 0.8632 | 0.8517 | 0.9108 | **0.8619** | **0.9526** |
| | Species | 0.1627 | 0.6045 | 0.2733 | 0.6597 | 0.3187 | 0.6536 | 0.3029 | 0.6743 | 0.3546 | 0.7399 | **1.0000** | **1.0000** |
| Fruit360 | Color | 0.2544 | 0.6054 | 0.3776 | 0.6791 | 0.4264 | 0.6868 | 0.4097 | 0.6841 | 0.4594 | 0.7392 | **0.6239** | **0.8243** |
| | Species | 0.2184 | 0.5805 | 0.2985 | 0.6176 | 0.4142 | 0.6984 | 0.3861 | 0.6732 | 0.5139 | 0.7430 | **0.5284** | **0.7582** |
| Card | Order | 0.0807 | 0.7805 | 0.0792 | 0.7128 | 0.1225 | 0.7313 | 0.1144 | 0.7658 | 0.1440 | 0.8267 | **0.3653** | **0.8587** |
| | Suits | 0.0497 | 0.3587 | 0.0430 | 0.3638 | 0.0676 | 0.3801 | 0.0716 | 0.3715 | 0.0873 | 0.4228 | **0.2734** | **0.7039** |
| CMUface | Emotion | 0.1284 | 0.6736 | 0.1433 | 0.5268 | 0.1592 | 0.6630 | 0.0422 | 0.5932 | 0.0161 | 0.5367 | **0.1786** | **0.7105** |
| | Glass | 0.1420 | 0.5745 | 0.1201 | 0.4905 | 0.1493 | 0.6209 | 0.1929 | 0.5627 | 0.1039 | 0.5361 | **0.3402** | **0.7068** |
| | Identity | 0.3892 | 0.7326 | 0.4637 | 0.6247 | 0.5607 | 0.7635 | 0.5109 | 0.8260 | 0.5875 | 0.8334 | **0.6625** | **0.9496** |
| | Pose | 0.3687 | 0.6322 | 0.3254 | 0.6028 | 0.2290 | 0.5029 | 0.4437 | 0.6114 | 0.1320 | 0.5517 | **0.4693** | **0.6624** |
| Stanford Cars | Color | 0.2331 | 0.6158 | 0.2103 | 0.5802 | 0.2465 | 0.6779 | 0.2336 | 0.6552 | 0.2736 | 0.7525 | **0.7360** | **0.9193** |
| | Type | 0.1325 | 0.5336 | 0.1650 | 0.5634 | 0.2063 | 0.6217 | 0.1963 | 0.5643 | 0.2364 | 0.7356 | **0.6355** | **0.8399** |
| Flowers | Color | 0.2561 | 0.5965 | 0.2938 | 0.5860 | 0.3329 | 0.6214 | 0.3169 | 0.6127 | 0.3556 | 0.6931 | **0.6426** | **0.7984** |
| | Species | 0.1326 | 0.5273 | 0.1561 | 0.6065 | 0.1894 | 0.6195 | 0.1887 | 0.6077 | 0.1996 | 0.6227 | **0.6013** | **0.8103** |

Table 2. Quantitative comparison. The significantly best results with 95% confidence are in bold.

# Resultados

## Multi-MAP vs Métodos existentes

| Dataset | Clustering | CLIP$_{GPT}$ NMI | CLIP$_{GPT}$ RI | CLIP$_{label}$ NMI | CLIP$_{label}$ RI | Multi-MaP NMI | Multi-MaP RI |
|---|---|---|---|---|---|---|---|
| ALOI | Color | 0.8581 | 0.9407 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | Shape | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Fruit | Color | 0.7912 | 0.9075 | **0.8629** | **0.9780** | 0.8619 | 0.9526 |
| | Species | 0.9793 | 0.9919 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Fruit360 | Color | 0.5613 | 0.7305 | 0.5746 | 0.7673 | **0.6239** | **0.8243** |
| | Species | 0.4370 | 0.7552 | **0.5364** | **0.7631** | 0.5284 | 0.7582 |
| Card | Order | 0.3518 | 0.8458 | 0.3518 | 0.8458 | **0.3653** | **0.8587** |
| | Suits | 0.2711 | 0.6123 | 0.2711 | 0.6123 | **0.2734** | **0.7039** |
| CMUface | Emotion | 0.1576 | 0.6532 | 0.1590 | 0.6619 | **0.1786** | **0.7105** |
| | Glass | 0.2905 | 0.6869 | **0.4686** | **0.7505** | 0.3402 | 0.7068 |
| | Identity | 0.1998 | 0.6388 | 0.2677 | 0.7545 | **0.6625** | **0.9496** |
| | Pose | 0.4088 | 0.6473 | 0.4691 | 0.6409 | **0.4693** | **0.6624** |
| Stanford Cars | Color | 0.6539 | 0.8237 | 0.6830 | 0.8642 | **0.7360** | **0.9193** |
| | Type | 0.6207 | 0.7931 | **0.6429** | **0.8456** | 0.6355 | 0.8399 |
| Flowers | Color | 0.5653 | 0.7629 | 0.5828 | 0.7836 | **0.6426** | **0.7984** |
| | Species | 0.5620 | 0.7553 | **0.6019** | 0.7996 | 0.6013 | **0.8103** |

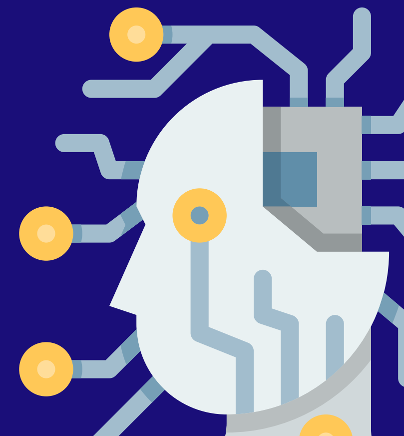
Table 3. Variants of CLIP. The significantly best results with 95% confidence are in bold.

# Resultados

## Estudio de ablación

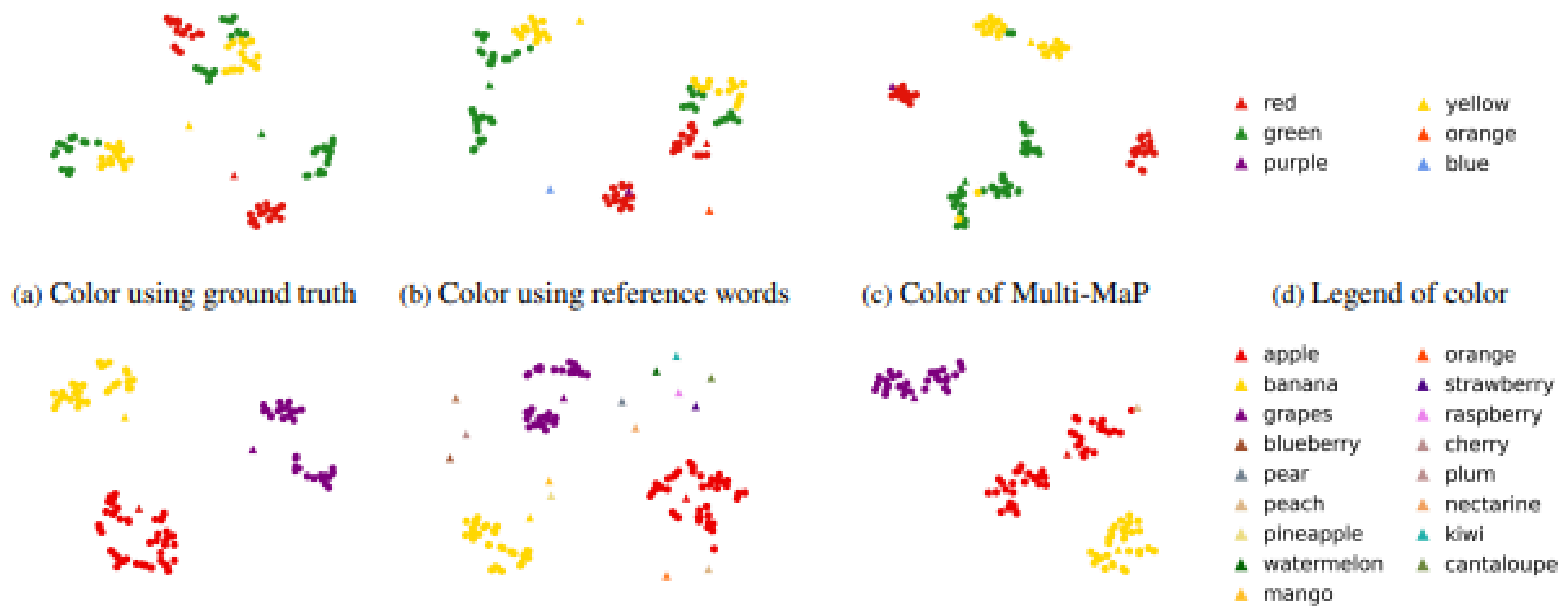
| | | Multi-MaP$_p$ | | Multi-MaP$_c$ | | Multi-MaP$_r$ | | Multi-MaP$_{cr}$ | | Multi-MaP | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Modules | Proxy Learning | ✓ | | ✓ | | ✓ | | ✓ | | ✓ | |
| | Concept Word | × | | ✓ | | × | | ✓ | | ✓ | |
| | Reference Word | × | | × | | ✓ | | ✓ | | ✓ | |
| | Contrastive Concepts | × | | × | | × | | × | | ✓ | |
| | | NMI↑ | RI↑ | NMI↑ | RI↑ | NMI↑ | RI↑ | NMI↑ | RI↑ | NMI↑ | RI↑ |
| ALOI [7] | Color | 0.9619 | 0.9826 | 1.0000 | 1.0000 | 0.9795 | 0.9869 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | Shape | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Fruit [12] | Color | 0.7642 | 0.8439 | 0.8215 | 0.9283 | 0.8136 | 0.9073 | 0.8484 | 0.9308 | **0.8619** | **0.9526** |
| | Species | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Fruit360 [35] | Color | 0.5643 | 0.7665 | 0.6217 | 0.7836 | 0.5910 | 0.7746 | 0.6089 | 0.7965 | **0.6239** | **0.8243** |
| | Species | 0.5077 | 0.7368 | 0.5137 | 0.7436 | 0.5094 | 0.7425 | 0.5199 | 0.7428 | **0.5284** | **0.7582** |
| Card [35] | Order | 0.1932 | 0.8152 | 0.3568 | 0.8472 | 0.3113 | 0.8229 | 0.3616 | 0.8094 | **0.3653** | **0.8587** |
| | Suits | 0.2375 | 0.6282 | 0.2696 | 0.6641 | 0.2498 | 0.6365 | 0.2562 | 0.6599 | **0.2734** | **0.7039** |
| CMUface [10] | Emotion | 0.1690 | 0.6170 | 0.1714 | 0.6229 | 0.1697 | 0.6360 | 0.1713 | 0.6843 | **0.1786** | **0.7105** |
| | Glass | 0.3112 | 0.6911 | 0.3269 | 0.7136 | 0.3162 | 0.6917 | 0.3370 | 0.7108 | **0.3402** | **0.7068** |
| | Identity | 0.5617 | 0.8234 | 0.6243 | 0.8359 | 0.5839 | 0.8263 | 0.6391 | 0.8946 | **0.6625** | **0.9496** |
| | Pose | 0.4361 | 0.6386 | 0.4550 | 0.6499 | 0.4381 | 0.6429 | 0.4387 | 0.6489 | **0.4693** | **0.6624** |
| Stanford cars [14] | Color | 0.5939 | 0.7835 | 0.6836 | 0.8659 | 0.6729 | 0.8638 | 0.7112 | 0.9117 | **0.7360** | **0.9193** |
| | Type | 0.5569 | 0.7996 | 0.6383 | 0.8271 | 0.6091 | 0.8046 | 0.6289 | 0.8181 | **0.6355** | **0.8399** |
| Flowers [20] | Color | 0.5783 | 0.7723 | 0.5830 | 0.7833 | 0.5987 | 0.7849 | 0.6216 | 0.7941 | **0.6426** | **0.7984** |
| | Species | 0.5704 | 0.7608 | 0.5744 | 0.7842 | 0.5723 | 0.7811 | 0.5846 | 0.7892 | **0.6013** | **0.8103** |

Table 4. Components ablation. All of our components boost performance consistently in all benchmark multi-clustering vision tasks.



# Resultados

## Visualización




(a) Color using ground truth (b) Color using reference words (c) Color of Multi-MaP (d) Legend of color

(e) Species using ground truth (f) Species using reference words (g) Species of Multi-MaP (h) Legend of species

Figure 4. Visualization of feature embeddings and related labels. The points represent the image or pseudo-word embeddings, and the triangles represent the prompt or label embeddings. Different colors represent different labels, which are indicated by the text next to the triangles.

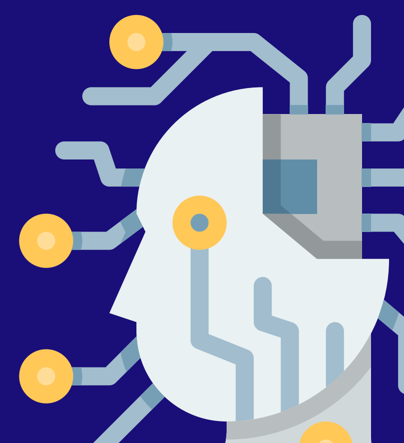
# Resultados

Desempeño destacado de Multi-MaP

- **Multi-MaP logra resultados superiores en todos los benchmarks públicos evaluados de clustering múltiple en visión por computadora.**

- **Demuestra por primera vez que CLIP puede distinguir diferentes aspectos semánticos de imágenes cuando se orienta adecuadamente por prompts específicos del usuario.**

- **La propuesta permite obtener agrupamientos personalizados, sin necesidad de etiquetas explícitas.**



# Resultados

## Limitación y futuro trabajo

- **Multi-MaP requiere que las etiquetas sean semánticamente interpretables para generar buenas referencias.**

- **Cuando las etiquetas no tienen interpretación semántica clara, Multi-MaP pierde eficacia al no poder generar referencias útiles automáticas desde GPT-4.**

# Resultados

## Limitación y futuro trabajo

- **Integración con WordNet y otros recursos externos**
- **Adaptación a dominios específicos**
- **Mejoras en selección automática de conceptos**

# Referencias


(1) ERIC BAE AND JAMES BAILEY. COALA: A NOVEL APPROACH FOR THE EXTRACTION OF AN ALTERNATE CLUSTERING OF HIGH QUALITY AND HIGH DISSIMILARITY. IN ICDM, PAGES 53–62. IEEE, 2006. 1, 2

(2) CHRISTOPHER M BISHOP AND NASSER M NASRABADI. PATTERN RECOGNITION AND MACHINE LEARNING. SPRINGER, 2006. 1

(3) STEPHEN P. BOYD AND LIEVEN VANDENBERGHE. CONVEX OPTIMIZATION. CAMBRIDGE UNIVERSITY PRESS, 2014. 4

(4) XUAN HONG DANG AND JAMES BAILEY. GENERATION OF ALTERNATIVE CLUSTERINGS USING THE CAMI APPROACH. IN PROCEEDINGS OF THE 2010 SIAM INTERNATIONAL CONFERENCE ON DATA MINING, PAGES 118–129. SIAM, 2010. 2

(5) CHRISTIANE FELLBAUM. WORDNET. IN THEORY AND APPLICATIONS OF ONTOLOGY: COMPUTER APPLICATIONS, PAGES 231–243. SPRINGER, 2010. 6

(6) PENG GAO, SHIJIE GENG, RENRUI ZHANG, TELI MA, RONGYAO FANG, YONGFENG ZHANG, HONGSHENG LI, AND YU QIAO. CLIP-ADAPTER: BETTER VISION-LANGUAGE MODELS WITH FEATURE ADAPTERS. INTERNATIONAL JOURNAL OF COMPUTER VISION, PAGES 1–15, 2023. 1, 3

(7) JAN-MARK GEUSEBROEK, GERTJAN J BURGHOUTS, AND ARNOLD WM SMEULDERS. THE AMSTERDAM LIBRARY OF OBJECT IMAGES. INTERNATIONAL JOURNAL OF COMPUTER VISION, 61:103–112, 2005. 5, 7

(8) DAVID GONDEK AND THOMAS HOFMANN. CONDITIONAL INFORMATION BOTTLENECK CLUSTERING. IN 3RD IEEE INTERNATIONAL CONFERENCE ON DATA MINING, WORKSHOP ON CLUSTERING LARGE DATA SETS, PAGES 36–42, 2003. 2

(9) JORIS GUERIN AND BYRON BOOTS. IMPROVING IMAGE CLUSTERING ´ WITH MULTIPLE PRETRAINED CNN FEATURE EXTRACTORS. IN BRITISH MACHINE VISION CONFERENCE 2018, BMVC 2018, 2018. 1, 2, 5

(10) STEPHAN GUNNEMANN, INES F ¨ ARBER, MATTHIAS R ¨ UDIGER, AND ¨ THOMAS SEIDL. SMVC: SEMI-SUPERVISED MULTI-VIEW CLUSTERING IN SUBSPACE PROJECTIONS. IN PROCEEDINGS OF THE 20TH ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, PAGES 253–262, 2014. 5, 6, 7

# Referencias


(11) JUHUA HU AND JIAN PEI. SUBSPACE MULTI-CLUSTERING: A REVIEW. KNOWLEDGE AND INFORMATION SYSTEMS, 56(2):257–284, 2018. 2
(12) JUHUA HU, QI QIAN, JIAN PEI, RONG JIN, AND SHENGHUO ZHU. FINDING MULTIPLE STABLE CLUSTERINGS. KNOWLEDGE AND INFORMATION SYSTEMS, 51(3):991–1021, 2017. 1, 2, 3, 5, 7
(13) TONY HUANG, JACK CHU, AND FANGYUN WEI. UNSUPERVISED PROMPT LEARNING FOR VISION-LANGUAGE MODELS. ARXIV PREPRINT ARXIV:2204.03649, 2022. 3
(14) JONATHAN KRAUSE, MICHAEL STARK, JIA DENG, AND LI FEI-FEI. 3D OBJECT REPRESENTATIONS FOR FINE-GRAINED CATEGORIZATION. IN PROCEEDINGS OF THE IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION WORKSHOPS, PAGES 554–561, 2013. 5, 7
(15) STUART LLOYD. LEAST SQUARES QUANTIZATION IN PCM. IEEE TRANSACTIONS ON INFORMATION THEORY, 28(2):129–137, 1982. 5
(16) JAMES MACQUEEN ET AL. SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS. IN PROCEEDINGS OF THE FIFTH BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY, PAGES 281–297. OAKLAND, CA, USA, 1967. 1
(17) SACHIT MENON AND CARL VONDRICK. VISUAL CLASSIFICATION VIA DESCRIPTION FROM LARGE LANGUAGE MODELS. ARXIV PREPRINT ARXIV:2210.07183, 2022. 3
(18) LUKAS MIKLAUTZ, DOMINIK MAUTZ, MUZAFFER CAN ALTINIGNELI, CHRISTIAN BOHM, AND CLAUDIA PLANT. DEEP EMBEDDED NON- " REDUNDANT CLUSTERING. IN PROCEE DINGS OF THE AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE, PAGES 5174–5181, 2020. 1, 2, 5
(19) ANDREW NG, MICHAEL JORDAN, AND YAIR WEISS. ON SPECTRAL CLUSTERING: ANALYSIS AND AN ALGORITHM. ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, 14, 2001. 1
(20) MARIA-ELENA NILSBACK AND ANDREW ZISSERMAN. AUTOMATED FLOWER CLASSIFICATION OVER A LARGE NUMBER OF CLASSES. IN 2008 SIXTH INDIAN CONFERENCE ON COMPUTER VISION, GRAPHICS & IMAGE PROCESSING, PAGES 722–729. IEEE, 2008. 5, 7

# Referencias


(21) ZIJIE QI AND IAN DAVIDSON. A PRINCIPLED AND FLEXIBLE FRAMEWORK FOR FINDING ALTERNATIVE CLUSTERINGS. IN SIGKDD, PAGES 717–726, 2009. 2

(22) QI QIAN, LEI SHANG, BAIGUI SUN, JUHUA HU, TACOMA TACOMA, HAO LI, AND RONG JIN. SOFTTRIPLE LOSS: DEEP METRIC LEARNING WITHOUT TRIPLET SAMPLING. IN ICCV, PAGES 6449–6457. IEEE, 2019. 3

(23) QI QIAN, YUANHONG XU, JUHUA HU, HAO LI, AND RONG JIN. UNSUPERVISED VISUAL REPRESENTATION LEARNING BY ONLINE CONSTRAINED K-MEANS. IN PROCEEDINGS OF THE IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, PAGES 16640–16649, 2022. 1

(24) QI QIAN, YUANHONG XU, AND JUHUA HU. INTRA-MODAL PROXY LEARNING FOR ZERO-SHOT VISUAL CATEGORIZATION WITH CLIP. IN THIRTY-SEVENTH CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS, NEURIPS 2023, 2023. 3

(25) ALEC RADFORD, JONG WOOK KIM, CHRIS HALLACY, ADITYA RAMESH, GABRIEL GOH, SANDHINI AGARWAL, GIRISH SASTRY, AMANDA ASKELL, PAMELA MISHKIN, JACK CLARK, ET AL. LEARNING TRANSFERABLE VISUAL MODELS FROM NATURAL LANGUAGE SUPERVISION. IN INTERNATIONAL CONFERENCE ON MACHINE LEARNING, PAGES 8748–8763. PMLR, 2021. 1, 2, 3

(26) WILLIAM M RAND. OBJECTIVE CRITERIA FOR THE EVALUATION OF CLUSTERING METHODS. JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, 66(336):846–850, 1971. 5

(27) LIANGRUI REN, GUOXIAN YU, JUN WANG, LEI LIU, CARLOTTA DOMENICONI, AND XIANGLIANG ZHANG. A DIVERSIFIED ATTENTION MODEL FOR INTERPRETABLE MULTIPLE CLUSTERINGS. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, 2022. 1, 2, 5

(28) ROBIN ROMBACH, ANDREAS BLATTMANN, DOMINIK LORENZ, PATRICK ESSER, AND BJORN OMMER. HIGH-RESOLUTION IMAGE SYN- ¨ THESIS WITH LATENT DIFFUSION MODELS. IN CVPR, PAGES 10674– 10685. IEEE, 2022. 3

(29) MANLI SHU, WEILI NIE, DE-AN HUANG, ZHIDING YU, TOM GOLDSTEIN, ANIMA ANANDKUMAR, AND CHAOWEI XIAO. TESTTIME PROMPT TUNING FOR ZERO-SHOT GENERALIZATION IN VISIONLANGUAGE MODELS. ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, 35:14274–14289, 2022. 3

(30) JUNYANG WANG, YUANHONG XU, JUHUA HU, MING YAN, JITAO SANG, AND QI QIAN. IMPROVED VISUAL FINE-TUNING WITH NATURAL LANGUAGE SUPERVISION. ARXIV PREPRINT ARXIV:2304.01489, 2023. 1, 3

# Referencias

(31) SHAOWEI WEI, JUN WANG, GUOXIAN YU, CARLOTTA DOMENICONI, AND XIANGLIANG ZHANG. MULTI-VIEW MULTIPLE CLUSTERINGS USING DEEP MATRIX FACTORIZATION. IN PROCEEDINGS OF THE AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE, PAGES 6348–6355, 2020. 2
(32) JV WHITE, SAM STEINGOLD, AND CG FOURNELLE. PERFORMANCE METRICS FOR GROUP-DETECTION ALGORITHMS. PROCEEDINGS OF INTERFACE, 2004, 2004. 5
(33) JUNYUAN XIE, ROSS GIRSHICK, AND ALI FARHADI. UNSUPERVISED DEEP EMBEDDING FOR CLUSTERING ANALYSIS. IN INTERNATIONAL CONFERENCE ON MACHINE LEARNING, PAGES 478–487, 2016. 1
(34) SEN YANG AND LIJUN ZHANG. NON-REDUNDANT MULTIPLE CLUSTERING BY NONNEGATIVE MATRIX FACTORIZATION. MACHINE LEARNING, 106(5):695–712, 2017. 2
(35) JIAWEI YAO, ENBEI LIU, MAHAM RASHID, AND JUHUA HU. AUGDMC: DATA AUGMENTATION GUIDED DEEP MULTIPLE CLUSTERING. IN INNS DLIA@IJCNN, 2023. 1, 3, 5, 7
(36) KAIYANG ZHOU, JINGKANG YANG, CHEN CHANGE LOY, AND ZIWEI LIU. LEARNING TO PROMPT FOR VISION-LANGUAGE MODELS. INTERNATIONAL JOURNAL OF COMPUTER VISION, 130(9):2337–2348, 2022. 3