Learning to Select Visual In-Context Demonstrations

Eugene Lee[1], Yu-Chi Lin[2], Jiajie Diao[1]
[1] University of Cincinnati [2] University of California, Los Angeles
eugene.lee@uc.edu, yclin0177@g.ucla.edu, jiajie.diao@uc.edu

Abstract

Multimodal Large Language Models (MLLMs) adapt to visual tasks via in-context learning (ICL), which relies heavily on demonstration quality. The dominant demonstration selection strategy is unsupervised k-Nearest Neighbor (kNN) search. While simple, this similarity-first approach is suboptimal for complex factual regression tasks; it selects redundant examples that fail to capture the task's full output range. We reframe selection as a sequential decision-making problem and introduce Learning to Select Demonstrations (LSD), training a Reinforcement Learning agent to construct optimal demonstration sets. Using a Dueling DQN with a query-centric Transformer Decoder, our agent learns a policy that maximizes MLLM downstream performance. Evaluating across five visual regression benchmarks, we uncover a crucial dichotomy: while kNN remains optimal for subjective preference tasks, LSD significantly outperforms baselines on objective, factual regression tasks. By balancing visual relevance with diversity, LSD better defines regression boundaries, illuminating when learned selection is strictly necessary for visual ICL.

1. Introduction

Multimodal Large Language Models (MLLMs) and Large Language Models (LLMs) have demonstrated remarkable abilities in complex tasks through in-context learning (ICL) [6], including mathematical reasoning [41]. This paradigm has driven a significant shift in few-shot learning (FSL). With the advent of powerful Vision Foundation Models (VFMs) and Vision-Language Models (VLMs), ICL is now the dominant approach for few-shot adaptation. Consequently, as [58] notes, the core research question has pivoted from training few-shot learners to effectively prompting massively pre-trained models.

Project page and code: https://eugenelet.github.io/LSD-Project/

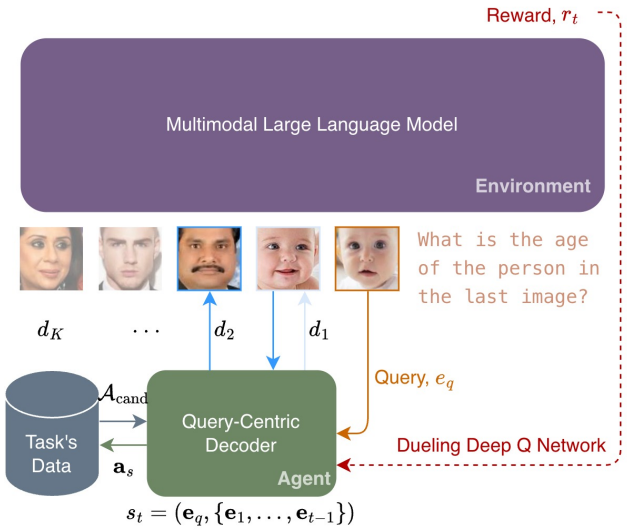


Figure 1. **An overview of our LSD (Learning to Select Demonstrations) framework.** The process is a training loop where the MLLM acts as the Environment. **(1)** The **Agent** (a Dueling DQN) receives the current state s_t, which contains the query embedding $\mathbf{e}_q$ and the embeddings of all previously selected demonstrations $\{\mathbf{e}_1, \ldots, \mathbf{e}_{t-1}\}$. **(2)** The agent's query-centric decoder outputs an advantage query $\mathbf{a}_s$, which is used to retrieve candidates A_{cand} from the **Task's Data** via FAISS. **(3)** The agent selects the next best demonstration, d_t. **(4)** The full prompt (including the selected demos $d_1 \ldots d_K$ and the query) is sent to the **MLLM** (Environment), which makes a prediction. **(5)** A **Reward** r_t is calculated based on the prediction's accuracy (e.g., MAE). **(6)** This reward is used to update the agent's policy.

However, ICL efficacy is highly sensitive to prompt configuration, especially the selection and ordering of demonstration examples [22, 38]. The impact of effective ICL spans diverse applications, including data engineering [5, 14, 37], model augmentation [30], knowledge updating [4], model safety [24, 26], and sentiment analysis [36, 44, 47, 52].

The most common selection strategy relies on unsupervised nearest neighbor (kNN) retrieval based on feature similarity [21, 29, 32]. While simple, this approach is of-

ten sub-optimal due to a lack of task-specific supervision [31, 38, 50, 53]. Its core "similarity-priority" assumption exhibits limited predictive power [42] and frequently yields redundant demonstration sets that provide misleading contextual information [16].

To move beyond simple similarity, research has explored *demonstration ordering*—arranging examples by proximity [21] or complexity [22]—and *demonstration construction*, emphasizing diversity [2] or using LLMs to generate new demonstrations [12, 15, 46, 48]. For visual ICL, complex retrieval-reranking paradigms have been proposed [57], alongside metrics designed to select for "representativeness" [11] or to explicitly model structural complexity [16].

A more fundamental critique, inspired by hard negative mining [17], argues these approaches over-index on positive, high-similarity examples. This has prompted a paradigm shift reframing shot selection as a sequential decision-making problem aimed at finding the most "informative" examples [23]. This view treats demonstration selection as a task for a Reinforcement Learning (RL) agent, learning a policy to maximize cumulative rewards tied to final ICL accuracy [53], shifting retrieval from simple visual similarity to a more abstract, task-oriented "reasoning process similarity" [29].

We embrace this sequential paradigm to address a critical gap in visual ICL: understanding *when* learned selection is actually necessary. Building on efforts utilizing LLM feedback [53, 56] or RL frameworks [39], we propose *LSD (Learning to Select Demonstrations)*, a novel RL framework that trains a Dueling DQN agent to sequentially construct demonstration sets for visual regression tasks. Our key hypothesis is that the optimal selection strategy depends fundamentally on whether the task is *objective* or *subjective*. For objective, factual tasks, the optimal set must contain diverse "boundary" examples that help the MLLM model the entire regression space. Conversely, for subjective preference tasks, a simple visual anchor often suffices. As shown in Fig. 1, our agent uses a query-centric Transformer Decoder to learn a policy that actively balances visual relevance with necessary diversity, avoiding the redundancy trap of kNN to maximize accuracy on complex objective domains.

Our main contributions are:

- We introduce LSD, a novel framework that successfully reframes K-shot demonstration selection as a sequential decision-making problem, scaling to dataset-level action spaces using a Dueling DQN agent and a query-centric Transformer Decoder.
- We conduct a comprehensive study on the efficacy of learned selection policies across five diverse visual regression benchmarks (UTKFace, AVA, SCUT-FBP5500, KonIQ-10k, and KADID-10k).
- We reveal a critical, task-dependent dichotomy in visual ICL: while unsupervised similarity search (kNN) remains highly effective for subjective preference tasks, our learned, diversity-aware policy is strictly necessary to achieve state-of-the-art performance on objective visual regression tasks.

2. Related Work

Our research builds upon a large body of work in visual in-context learning (ICL), particularly on the critical problem of demonstration selection.

Demonstration Selection for In-Context Learning. The performance of ICL is known to be highly sensitive to the choice of demonstration examples [22, 54]. This has been shown in various domains, with recent work demonstrating that MLLMs like GPT-4V can classify specialized medical images (e.g., histopathology) with high accuracy using just a few well-chosen examples [10]. This sensitivity has spurred significant research into methods that move beyond simple kNN retrieval.

A primary challenge is selecting an optimal *set* of demonstrations, not just individual relevant examples. One line of work treats this as a subset selection problem. Yang *et al*. [48] proposed selecting a single, representative set of demonstrations applicable to all test instances, using a Determinantal Point Process (DPP) to ensure both quality and diversity. This task-level approach contrasts with our instance-level policy. Purohit *et al*. [28] (CASE) framed set selection as a multi-armed bandit problem, treating each subset as an "arm" and using a novel sampling strategy to efficiently find the best set while minimizing expensive LLM calls.

Another line of research explores the trade-off between the two main criteria for selection: similarity and diversity. While similarity-based retrieval is effective for simple tasks, Xiao *et al*. [43] systematically demonstrated that incorporating diversity is crucial for improving performance and robustness on complex tasks, such as math and code generation. This finding directly supports our hypothesis that a learned agent is necessary to intelligently balance these two competing objectives.

Rather than retrieving a static set, other methods treat selection as a sequential construction problem. Li [18] introduced SabER, a lightweight decoder that autoregressively selects *and* orders examples to construct an optimal prompt. While holistic, this approach is trained on scores from the target MLLM, making the resulting selector model-specific and requiring retraining for different MLLMs. Our RL-based approach, while also sequential, learns from a more generalizable reward signal (downstream MAE) and is not as tightly coupled to the reward model's architecture.

Finally, some work has focused on improving the retriever itself. Zhang *et al*. [54] proposed a supervised, contrastive learning framework to train a retriever that automatically selects examples which maximize downstream task performance. This highlights the value of task-specific supervision, which our RL framework incorporates via its reward function, in contrast to unsupervised similarity metrics.

Related ICL Training and Prompting Strategies. Beyond demonstration selection, other methods aim to improve IDCL by modifying the model's training or the prompting method itself. To better leverage few-shot examples, Lin *et al*. [8] introduced an "any-shot" training paradigm, showing that explicitly training models on ICL-formatted, multi-turn conversations enhances their ability to learn from context.

Diverging from selection entirely, other research explores how to elicit better reasoning from the LLM with no demonstrations. Yao [49], for instance, introduced Contrastive Prompting, a zero-shot method that instructs an LLM to generate both a correct and an incorrect solution. This process of explicit contrastive reasoning was shown to significantly boost performance on complex tasks by helping the model better discern the correct problem-solving path. Our work draws on a similar intuition: providing a diverse or "contrastive" set of demonstrations (e.g., high and low scores) in the prompt can serve a similar purpose, helping the model to "triangulate" the correct answer.

3. Method

The problem of selecting an optimal set of K demonstrations for in-context learning (ICL) can be framed as a sequential decision-making task. While unsupervised methods based on feature similarity are common [11, 21], they are often sub-optimal as they lack task-specific supervision [31, 38]. To overcome this, we adopt a Reinforcement Learning (RL) framework, similar to approaches in [39, 53], to learn a policy that iteratively constructs a high-quality demonstration set.

Our core contribution is a novel Dueling Deep Q-Network (DQN) [40] architecture specifically designed to handle the massive, discrete action space inherent in demonstration selection, where any sample from the entire dataset N can be chosen. Instead of a linear output layer of size N, our network computes Q-values by projecting the state representation into a query vector, which then interacts with the embedding of all possible actions via an efficient, approximate nearest-neighbor search.

3.1. Problem Formulation as an MDP

We model the K-shot demonstration selection process as a finite-horizon Markov Decision Process (MDP), defined by the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$:

- **State ($s_t \in \mathcal{S}$):** A state at step t (for $t = 1, \dots, K$) is defined by the query q and the ordered set of demonstrations selected so far, $D_{t-1} = \{d_1, \dots, d_{t-1}\}$. The initial state s_1 contains the query and one "anchor" demonstration found via nearest-neighbor search, to provide initial context.
- **Action ($a_t \in \mathcal{A}$):** An action is the selection of a new demonstration d_t from the pool of all available samples $\mathcal{C}$, excluding the query and any previously selected demonstrations: $a_t \in \mathcal{C} \setminus (\{q\} \cup D_{t-1})$. The action space $|\mathcal{A}|$ is thus $O(N)$, where N is the total number of samples in the dataset.
- **Transition ($\mathcal{P}$):** The state transition is deterministic. Upon taking action $a_t = d_t$ in state $s_t = (q, D_{t-1})$, the environment transitions to state $s_{t+1} = (q, D_t)$, where $D_t = D_{t-1} \cup \{d_t\}$. The episode terminates when K demonstrations have been selected ($t = K$).
- **Reward ($\mathcal{R}$):** The reward function is designed to optimize the marginal utility of each added demonstration. We define a MLLM scoring function, $R(s_t) = -\text{MAE}(\mathcal{V}(q, D_{t-1}))$, which queries the MLLM with the query q and demonstrations D_{t-1} and returns the negative Mean Absolute Error (MAE) of its prediction. The reward r_t for selecting action a_t is the *improvement* in this score:

$$r(s_t, a_t) = R(s_{t+1}) - R(s_t) \tag{1}$$

  This sparse reward encourages the agent to select samples that progressively refine the MLLM's accuracy. A large penalty is given for invalid actions (e.g., re-selecting a sample) or MLLM failures.
- **Discount (γ):** We use a discount factor γ to balance immediate and future rewards.

The agent's goal is to learn the optimal action-value function $Q^*(s, a)$, which represents the maximum expected cumulative reward $G_t = \sum_{i=t}^{K} \gamma^{i-t} r_i$ from state s by taking action a and following the optimal policy thereafter.

3.2. Dueling Q-Network for Large Action Spaces

A standard DQN is infeasible due to the $O(N)$ action space. We therefore employ a Dueling Q-Network architecture that leverages the underlying embedding space $\mathbb{R}^D$ of the samples. All N samples are represented by a D-dimensional embedding $\mathbf{e}_i$, pre-computed using a SigLIP model [51].

The $Q(s, a)$ function is decomposed into a state-value $V(s)$ and an action-advantage $A(s, a)$ [40]:

$$Q(s, a) = V(s) + \left(A(s, a) - \frac{1}{|\mathcal{A}|} \sum_{a' \in \mathcal{A}} A(s, a') \right) \tag{2}$$

Our network (Fig. 1) does not compute $A(s, a)$ for all $a \in \mathcal{A}$. Instead, it computes $V(s)$ and a D-dimensional

"advantage query" vector $\mathbf{a}_s$. The advantage $A(s, a_i)$ for a specific action (sample i) is then calculated as the inner product of this query with the action's embedding $\mathbf{e}_i$:

$$A(s, a_i) = \mathbf{a}_s^\top \mathbf{e}_i \tag{3}$$

This formulation assumes both $\mathbf{a}_s$ and $\mathbf{e}_i$ are L2-normalized, making the advantage a measure of cosine similarity.

3.3. Network Architecture

Our network consists of a query-centric state encoder and two dueling heads.

3.3.1. Query-Centric State Encoder

To produce a holistic state representation, we must fuse the query embedding $\mathbf{e}_q$ with the set of $t-1$ selected demonstration embeddings $\mathbf{E}_D = \{\mathbf{e}_1, \dots, \mathbf{e}_{t-1}\}$.

A common approach might be to simply concatenate these embeddings, $[\mathbf{e}_q; \mathbf{e}_1; \dots; \mathbf{e}_{t-1}]$, and pass them through a Transformer Encoder (i.e., using only self-attention). However, our initial experiments revealed a significant failure mode with this design: the agent was prone to *policy collapse*, learning to select a single, query-agnostic set of "generally good" demonstrations, regardless of the query's specific features. This indicates the self-attention mechanism failed to adequately prioritize the query's relationship to the demonstrations.

To solve this, we designed a *Query-Centric State Encoder* using a standard Transformer Decoder architecture [35] in a specific way. We feed the L2-normalized query embedding $\mathbf{e}_q$ as the *target* sequence (with a sequence length of 1) and the set of $t-1$ demonstration embeddings $\mathbf{E}_D$ as the *memory* sequence. As ICL is sensitive to demonstration order [21, 22], the demonstration embeddings are first augmented with a learned positional encoding $\mathbf{P} \in \mathbb{R}^{K \times D}$.

A standard Transformer Decoder layer contains three sub-layers: masked self-attention, cross-attention, and a feedforward network (FFN). Because our *target* sequence has a length of one, the initial *masked self-attention sub-layer is definitionally bypassed* (it's a no-op).

Therefore, the computation within each of the L decoder layers reduces to two critical steps:

1. *Cross-Attention:* The query representation $\mathbf{x}_q^{(l-1)}$ (from the previous layer) is used to generate the *Query (Q)* vector. This $\mathbf{Q}$ probes the *memory* (demos) $\mathbf{M} = \mathbf{E}_D + \mathbf{P}$, which provides the *Key (K)* and *Value (V)* vectors. This step computes an attention-weighted vector that represents the query contextualized by the demonstrations.
2. *Feedforward Network (FFN):* The resulting vector is then processed by a standard position-wise FFN to produce the layer's output, $\mathbf{x}_q^{(l)}$.

This process repeats for L layers, progressively refining the query embedding based on the provided demonstration context. The final output $\mathbf{c}_s$ is the fully contextualized query vector, which is then passed to the dueling heads. This computation, performed by the L-layer TransformerDecoder, is defined as:

$$\begin{aligned}
&\mathbf{c}_s = \text{TransformerDecoder}(\text{target} = \mathbf{e}_q, \text{memory} = \mathbf{M}) \\
&\mathbf{M} = \mathbf{E}_D + \mathbf{P}\,; \quad \mathbf{x}_q^{(0)} = \mathbf{e}_q \\
&\mathbf{x}'_q = \mathbf{x}_q^{(l-1)} + \text{Softmax}\left(\frac{(\mathbf{x}_q^{(l-1)} W_Q^{(l)})(\mathbf{M} W_K^{(l)})^\top}{\sqrt{d_k}}\right)(\mathbf{M} W_V^{(l)}) \\
&\mathbf{x}_q^{(l)} = \mathbf{x}'_q + \text{FFN}^{(l)}(\mathbf{x}'_q) \quad \forall l \in \{1, \dots, L\} \\
&\mathbf{c}_s = \mathbf{x}_q^{(L)}
\end{aligned} \tag{4}$$

where $\text{FFN}^{(l)}$ is the feedforward network for layer l. This design ensures the state representation $\mathbf{c}_s$ is always conditioned on the specific query, mitigating policy collapse and enabling the agent to learn a query-specific selection policy.

3.3.2. Dueling Heads

The context vector $\mathbf{c}_s$ is passed to two separate heads:

1. **Value Head:** A simple linear layer that estimates the state-value $V(s)$:

$$V(s) = \mathbf{w}_v^\top \mathbf{c}_s + b_v \tag{5}$$

   where $\mathbf{w}_v \in \mathbb{R}^D$ and b_v is a scalar bias.
2. **Advantage Head:** A linear layer followed by L2 normalization, which produces the D-dimensional advantage query vector $\mathbf{a}_s$:

$$\mathbf{a}_s = \frac{\mathbf{W}_a \mathbf{c}_s + \mathbf{b}_a}{\|\mathbf{W}_a \mathbf{c}_s + \mathbf{b}_a\|_2} \tag{6}$$

   where $\mathbf{W}_a \in \mathbb{R}^{D \times D}$ and $\mathbf{b}_a \in \mathbb{R}^D$.

3.4. Approximate Q-Learning for Large Action Spaces

The Q-learning update requires computing the target value y_t, which depends on $\max_{a'} Q(s', a')$. Finding the true maximum would require N dot products (Eq. (3)), which is computationally prohibitive.

To solve this, we leverage Approximate Nearest Neighbor (ANN) search. We build a FAISS (IVFPQ) index [7] on the embeddings $\{\mathbf{e}_i\}_{i=1}^N$ of all dataset samples. This index can efficiently retrieve the $\mathcal{N}$ candidate actions whose embeddings have the highest inner product with a given advantage query vector $\mathbf{a}_s$.

3.4.1. Action Selection

We use an ϵ-greedy policy. With probability ϵ, we explore by selecting a random valid action from the $\mathcal{N}$ candidates returned by FAISS. With probability $1-\epsilon$, we exploit by executing the following steps:

1. Compute the state-value $V(s_t)$ and the advantage query $\mathbf{a}_{s_t}$ using the policy network Q_θ.
2. Use the FAISS index to retrieve the top $\mathcal{N}$ candidate actions: $\mathcal{A}_{\text{cand}} = \text{FAISS}(\mathbf{a}_{s_t}, \mathcal{N})$.
3. Calculate the advantage $A(s_t, a_j)$ for all $a_j \in \mathcal{A}_{\text{cand}}$ using Eq. (3).
4. Approximate the mean advantage using only these candidates: $\bar{A} \approx \frac{1}{\mathcal{N}} \sum_{a_j \in \mathcal{A}_{\text{cand}}} A(s_t, a_j)$.
5. Select the best action a_t according to the dueling Q-value (Eq. (2)):

$$a_t = \text{argmax}_{a_j \in \mathcal{A}_{\text{cand}}} \left(V(s_t) + (A(s_t, a_j) - \bar{A}) \right)$$

3.4.2. Optimization

We store transitions $(s_t, a_t, r_t, s_{t+1}, \text{done})$ in a replay buffer $\mathcal{B}$. For a mini-batch of B transitions, we compute the target y_t using the target network Q_{θ^-}:

$$y_t = r_t + \gamma(1 - \text{done}) \cdot \max_{a' \in \mathcal{A}'_{\text{cand}}} Q(s_{t+1}, a'; \theta^-) \quad (7)$$

where the max operation is performed efficiently using the same FAISS-based approximation on the target network's advantage query $\mathbf{a}_{s'}$.

The policy network Q_θ is then updated by minimizing the Smooth L1 (Huber) Loss between the predicted $Q(s_t, a_t; \theta)$ and the target y_t:

$$L(\theta) = \frac{1}{B} \sum_{(s,a,r,s') \in \mathcal{B}} \mathcal{L}_{\text{Huber}} \left(y_t - Q(s_t, a_t; \theta) \right) \quad (8)$$

The target network weights θ^- are updated via a soft polyak average: $\theta^- \leftarrow \tau\theta + (1 - \tau)\theta^-$.

4. Experiments

We conduct a comprehensive set of experiments to evaluate the effectiveness of our proposed demonstration selection method, which we refer to as *LSD* (Learning to Select Demonstrations). Our evaluation is designed to answer several key questions:

1. Does our method outperform standard unsupervised (kNN) and random selection baselines in terms of downstream task performance?
2. How does the performance scale with the number of demonstrations (K)?
3. Does our agent learn a selection policy that is qualitatively different from the baselines (e.g., by balancing relevance and diversity)?
4. Can a policy learned using reward signals from one MLLM generalize to improve the performance of other, unseen MLLMs?

To answer these, we evaluate on a diverse set of challenging visual regression tasks and compare against strong baselines.

4.1. Datasets

We focus on visual regression tasks, as they require nuanced reasoning from the MLLM that is often highly sensitive to demonstration quality. We use five public benchmark datasets:

- **UTKFace:** A large-scale face dataset with over 20,000 images, annotated with age, gender, and ethnicity. For our experiments, we use the *age prediction* task, which features a wide regression range from 0 to 116 years [55].
- **AVA (Aesthetic Visual Analysis):** A large-scale database of over 250,000 images, annotated with aesthetic scores (a regression task on a 1-10 scale), as well as semantic labels and photographic styles [25].
- **SCUT-FBP5500:** A facial beauty perception dataset consisting of 5,500 images of both Asian and Caucasian faces. Each image is annotated with an *attractiveness rating* on a 1-to-5 scale, providing a fine-grained regression task [19].
- **KonIQ-10k & KADID-10k:** Two large-scale Image Quality Assessment (IQA) datasets. KonIQ-10k contains 10,073 images with quality scores obtained via crowdsourcing, reflecting "authentic" perceptual quality [13]. KADID-10k contains 10,000 images generated from 81 pristine images, each distorted by 25 different degradation types at 5 levels, providing a benchmark for "synthetic" distortion [20].

4.2. Baselines

We compare the performance of our method, **LSD**, against two standard and widely-used demonstration selection baselines:

- **k-Nearest Neighbors (kNN):** This is the most common unsupervised baseline, based on the method in [21]. For a given query, we compute its SigLIP embedding and select the K samples from the training pool with the highest cosine similarity to the query embedding.
- **Random:** We randomly select K demonstrations from the training pool, excluding the query itself. This baseline helps establish whether the task benefits from ICL at all.
- **0-Shot:** We also report the performance of the MLLM with no demonstrations, which serves as the absolute performance floor and quantifies the overall benefit of ICL.

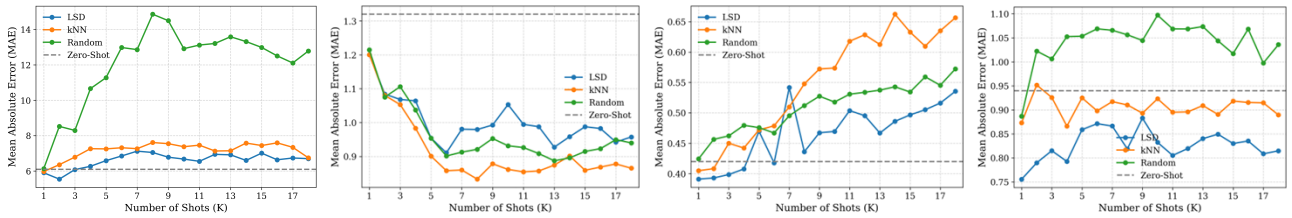
4.3. Implementation Details

MLLMs. Our experiments utilize three publicly available Multimodal Large Language Models: *Gemma 3 4B-it* [33], *Qwen 2.5 7B* [3], and *Phi-3.5-vision (4.2B)* [1]. Unless otherwise specified, our LSD agent is trained using reward signals generated by Gemma 3 4B-it, as described in Sec. 3.

Embeddings. All sample embeddings for both our method and the kNN baseline are $D = 768$ dimensional vectors extracted from the *SigLIP-base-patch16-224* vision model.

Table 1. **Main Performance (MAE ↓) Comparison vs. Number of Shots (K).** We report the Mean Absolute Error (MAE) for all methods on the five benchmark datasets, evaluated with Gemma 3 4B-it for $K \in \{1, 4, 8, 16\}$. Our proposed method, **LSD**, consistently outperforms all baselines, and the performance gap widens as K increases. The 0-shot and a fully **Supervised** (Sup.) baseline are also provided for reference. Best results are in **bold**.

			Random				kNN				LSD (Ours)			
Dataset	**Sup.**	**0-Shot**	K=1	K=4	K=8	K=16	K=1	K=4	K=8	K=16	K=1	K=4	K=8	K=16
UTKFace	4.42 [27]	6.10	6.14	10.66	14.86	12.51	5.98	7.27	7.61	7.60	**5.90**	**6.27**	**7.05**	**6.64**
AVA	-	1.32	1.21	1.03	0.92	0.92	**1.20**	**0.98**	**0.83**	**0.86**	**1.20**	1.06	0.98	0.98
SCUT-FBP5500	0.26 [45]	0.59	0.58	0.64	0.64	0.68	**0.53**	**0.39**	**0.40**	**0.44**	0.55	0.62	0.67	0.75
KonIQ-10k	0.39 [34]	0.42	0.42	0.48	**0.44**	0.56	0.40	0.44	0.55	0.61	**0.39**	**0.40**	0.51	**0.51**
KADID-10k	-	0.94	0.89	1.07	1.05	1.07	0.87	0.87	0.91	0.92	**0.76**	**0.79**	**0.82**	**0.84**



(a) UTKFace (Age Prediction) (b) AVA (Aesthetic Rating) (c) KonIQ-10k (Image Quality) (d) KADID-10k (Image Quality)

Figure 2. **Performance vs. Number of Shots (K) on four datasets.** We plot the MAE as K increases. The results are task-dependent: **(a), (c), (d) Objective Tasks (UTKFace, KonIQ, KADID):** Our LSD policy (blue) consistently outperforms the kNN baseline (orange). **(b) Subjective Task (AVA):** The kNN baseline, which is based on visual similarity, consistently outperforms LSD.

LSD Agent. Our Dueling DQN agent's state encoder is a Transformer Decoder with $L = 2$ layers and $H = 4$ attention heads. We use a discount factor $\gamma = 0.99$, a learning rate of 5×10^{-6}, a replay buffer of 50,000 transitions, and a batch size of 32. For efficient action selection (Eq. (7)), we use a FAISS (IVFPQ) index to retrieve $\mathcal{N} = 200$ candidates at each step. The agent is trained for 16,000 steps, which takes approximately 7 hours on a single NVIDIA A100 GPU.

4.4. Evaluation Protocols

We design several experiments to rigorously evaluate our method. While benchmarks like AVA, SCUT-FB5500, KonIQ-10k, and KADID-10k often report the Pearson Linear Correlation Coefficient (PLCC) or Spearman Rank Correlation Coefficient (SRCC), these metrics primarily measure the *monotonicity* or *correlation* of predictions against ground truth labels.

For our experiments, the primary metric is *Mean Absolute Error (MAE)*. This choice is critical as it directly aligns with our method's optimization objective. Our RL agent's reward signal is a function of the MLLM's prediction error (e.g., $r_t \propto -\text{MAE}$). Therefore, evaluating with MAE is the most direct and accurate measure of our agent's success, as it quantifies exactly what the policy was trained to improve: the absolute accuracy of the MLLM's prediction.

4.4.1. Main Performance vs. K

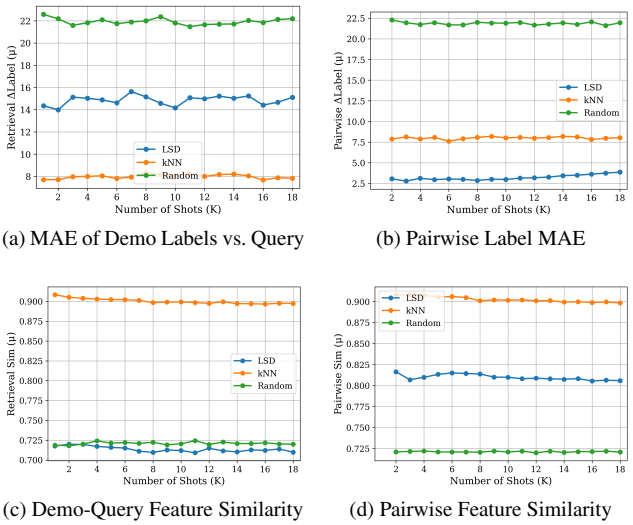
Our primary experiment evaluates MAE as a function of the number of demonstrations K. The full results are presented in Tab. 1, with performance scaling on key datasets shown in Fig. 2. The results reveal a clear, task-dependent pattern.

On the *objective* regression tasks—age prediction (UTKFace), and image quality assessment (KonIQ-10k and KADID-10k)—our learned LSD policy consistently and significantly outperforms the kNN baseline. As shown in Tab. 1, this performance gap is evident across $K = 4, 8,$ and 16 for UTKFace and across all K values for the IQA tasks. This highlights the efficiency of our learned, diversity-aware policy for tasks with a clear, factual ground truth.

Conversely, on the *subjective* tasks that rely on human judgment, such as aesthetic rating (AVA) and attractiveness rating (SCUT-FBP5500), the kNN baseline provides superior performance. This suggests that for these tasks, simple visual similarity (which kNN excels at) is a more effective strategy than our agent's learned policy.

4.4.2. Demonstration Set Analysis

To understand our agent's policy, we analyzed the selected demo sets on UTKFace (Fig. 3). Our analysis shows kNN uses a fixed, myopic strategy (visual similarity), while LSD learns a sophisticated policy by optimizing for MLLM performance.



(a) MAE of Demo Labels vs. Query

(b) Pairwise Label MAE

(c) Demo-Query Feature Similarity

(d) Pairwise Feature Similarity

Figure 3. **Demonstration Set Analysis on UTKFace, plotted against K shots.** (a) *MAE of Demo Labels vs. Query:* The MAE between selected demo labels and the query's true label. LSD finds demos with closer labels. (b) *Pairwise Label MAE:* The MAE computed over all pairwise label differences among the selected demos. (c) *Demo-Query Feature Similarity:* The cosine similarity between demo embeddings and the query embedding. LSD balances similarity with other factors. (d) *Pairwise Feature Similarity:* The cosine similarity between every pair of selected demonstrations. LSD actively seeks diverse (low-similarity) demos.

- **Visual Relevance vs. Diversity:** kNN selects highly redundant, visually similar demos. In contrast, LSD learns a more nuanced policy: it prioritizes relevance (Fig. 3(c)) but also actively seeks *diversity* by selecting new demos that are visually *dissimilar* from those already in the context (Fig. 3(d)).
- **Emergent Label-Awareness:** Most strikingly, Fig. 3(a) shows that by optimizing for the final reward, LSD *implicitly learns* to select demos that are closer in label-space to the query (lower MAE), despite its state containing no label information.

In short, LSD learns a superior policy that balances visual relevance with active diversity, resulting in an emergent strategy highly correlated with the task's underlying label structure.

4.4.3. Qualitative Analysis

Qualitative examples in Fig. 4 illustrate the core policy differences. The *kNN* baseline is myopic, invariably selecting a visually homogeneous and redundant set. For the 8-year-old query, it selects only other visually similar children. For the KADID-10k query, its policy is even more redundant, selecting only other distorted versions of the *same source image*.

In contrast, our *LSD* agent learns a sophisticated, context-building policy. For the UTKFace query, it selects a diverse spectrum of visual features, providing varied ages and appearances to help the MLLM understand the concept of "age". For the KADID-10k query, it learns to select crucial "boundary" examples, such as the pristine original (a high-score anchor) and images with entirely different distortion types from *different source images*. This diverse context better defines the entire regression space and drives LSD's superior performance on objective tasks (Tab. 1). However, this behavior also reveals a key limitation: for subjective preference tasks (e.g., AVA), this learned diversity introduces unnecessary variance, explaining why kNN's strict similarity approach remains superior.

4.4.4. Cross-MLLM Generalization

This experiment tests whether the learned policy is MLLM-agnostic. We take the single LSD agent trained using rewards from Gemma 3 4B-it and use this *frozen policy* to select demonstrations for *Qwen 2.5 7B* and *Phi-3.5-vision*. We evaluate performance on the objective UTKFace task, with results shown in Fig. 5.

The results show that our learned policy successfully transfers and remains highly effective, significantly outperforming the Random baseline on both models. The comparison to the strong kNN baseline is more nuanced. As shown in Fig. 5(a), our policy (blue line) maintains its performance advantage and consistently outperforms kNN (orange line) on Qwen 2.5 7B. On Phi-3.5-vision, shown in Fig. 5(b), our policy's performance is on par with kNN. This strongly suggests that our agent has learned a "fundamental" and generalizable policy that is not overfit to the original Gemma reward model, as it performs comparably or better than the strong kNN baseline on entirely unseen MLLMs.

Table 2. **Analysis of Selection Order (MAE ↓) at $K = 8$.** We compare the performance of the agent's learned demonstration sequence against the exact same set of demonstrations in a random order.

Dataset	LSD (Learned Order)	LSD (Shuffled Set)
UTKFace	7.05	**6.51**
AVA	**0.98**	1.04
SCUT-FBP5500	0.67	**0.53**
KonIQ-10k	**0.51**	0.59
KADID-10k	0.82	0.82

4.4.5. Analysis of Selection Order

Our agent selects demonstrations $d_1, \ldots, d_K$ sequentially. We sought to determine if this learned *order* is a critical part of its policy, or if the agent is primarily learning to select a good *set* of demonstrations. To test this, we conduct a permutation test. For each query, we first use our trained LSD agent to select its optimal $K = 8$ demonstrations and

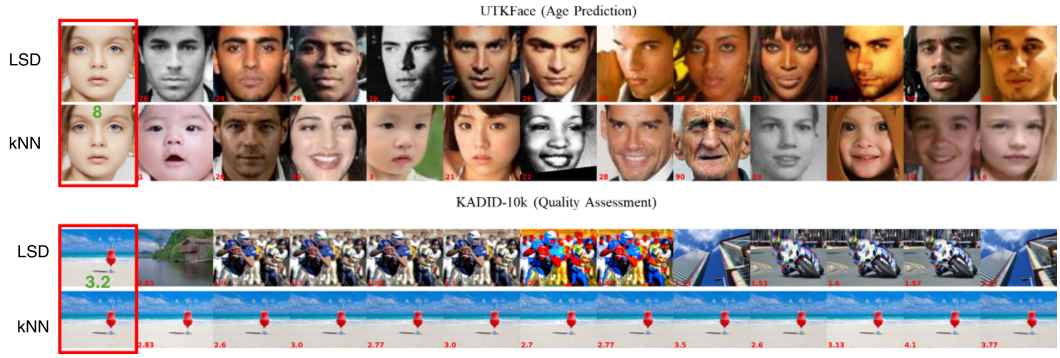
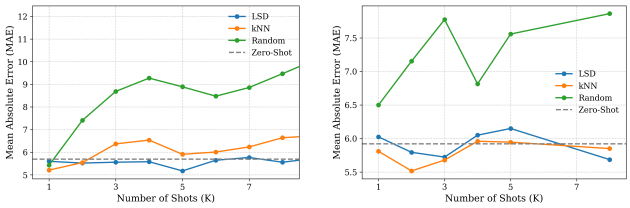


Figure 4. **Qualitative Comparison of Selected Demonstrations ($K = 12$).** (a) **UTKFace:** For an 8-year-old query, kNN selects only images with highly similar features (e.g., other young children). LSD selects a diverse spectrum of visual features (e.g., varied ages, genders, and lighting conditions) to build a richer context. (b) **KADID-10k:** For a motion-blurred query, kNN selects only other distorted versions of the *same source image*. LSD selects a varied set, including the pristine original and images with *different distortion types* from *different source images*, defining the quality boundaries.



(a) UTKFace on Qwen 2.5 7B (b) UTKFace on Phi-3.5-vision

Figure 5. **Cross-MLLM Generalization (MAE ↓) on UTKFace vs. Number of Shots (K).** We use the single LSD policy (trained on Gemma 3 4B-it) to select demos for two unseen MLLMs. The plots show our policy (blue line) versus the kNN (orange line) and Random (green line) baselines. **(a)** On Qwen 2.5 7B, our policy consistently outperforms kNN. **(b)** On Phi-3.5-vision, our policy performs on par with kNN. Both LSD and kNN significantly outperform the Random baseline.

record the MAE. Then, we randomly shuffle the order of those same 8 demonstrations and re-run inference.

As shown in Tab. 2, the 'Shuffled Set' performance is nearly identical to, and not consistently worse than, the agent's 'Learned Order'. This strongly suggests that the primary skill our agent has learned is the selection of an optimal *set* of demonstrations. The MLLM, in this case, appears robust to the permutation of those demonstrations, as long as the high-quality set is provided in the context.

4.4.6. Ablation Study: State Encoder Architecture

Our ablation study (Tab. 3) compared our *Query-Centric* model to a *Concat Input* baseline, which concatenates all embeddings ($[\mathbf{e}_q; \mathbf{E}_{t-1}]$) as a single 'tgt' sequence. The baseline exhibited a critical behavioral failure: *policy collapse*, learning to select the same non-query-specific demonstrations for all queries. This fundamental failure confirms its inferiority, despite its inconsistent MAE scores.

Table 3. **Ablation Study on Decoder Input Strategy (MAE ↓).** We compare our query-centric model against a standard decoder-only model (Concat Input) on UTKFace for $K \in \{4, 8, 16\}$, both using $L = 2$ layers. We also note the qualitative policy behavior.

	MAE ↓			
Decoder Input Strategy	K=4	K=8	K=16	**Policy Behavior**
Query-Centric	**6.27**	7.05	**6.64**	Query-specific demos
Concat Input	7.01	**6.42**	7.74	Non-query-specific

Our *Query-Centric* model successfully learned a query-specific policy with strong and stable performance, proving our architectural choice is essential to learn an effective, non-degenerate policy.

5. Conclusion

We introduced LSD, a novel framework that reframes in-context demonstration selection as a sequential decision-making problem. Powered by a query-centric Transformer Decoder, our Dueling DQN agent learns a selection policy by optimizing for downstream MLLM performance, scaling to massive $O(N)$ action spaces via efficient FAISS-based retrieval. Crucially, our comprehensive evaluation reveals a fundamental task-dependent dichotomy in visual ICL: while simple kNN retrieval remains highly effective for subjective preference tasks, our learned policy is strictly necessary to achieve superior performance on objective visual regression tasks. By actively balancing visual relevance with necessary diversity, LSD develops an emergent awareness of the label structure to better define regression boundaries. This non-degenerate, generalizable policy demonstrates a clear path forward, illuminating exactly when learning—rather than simply retrieving—is essential for optimal in-context demonstrations.

Acknowledgements

This work was supported by the National Institutes of Health (NIH R35GM128837).

References

[1] Marah Abdin, Jyoti Aneja, Harkirat Behl, Sébastien Bubeck, Ronen Eldan, Suriya Gunasekar, Michael Harrison, Russell J Hewett, Mojan Javaheripi, Piero Kauffmann, et al. Phi-4 technical report. *arXiv preprint arXiv:2412.08905*, 2024. 5

[2] Shengnan An, Zeqi Lin, Qiang Fu, Bei Chen, Nanning Zheng, Jian-Guang Lou, and Dongmei Zhang. How do in-context examples affect compositional generalization? *arXiv preprint arXiv:2305.04835*, 2023. 2

[3] Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, et al. Qwen2. 5-vl technical report. *arXiv preprint arXiv:2502.13923*, 2025. 5

[4] Nicola De Cao, Wilker Aziz, and Ivan Titov. Editing factual knowledge in language models. *arXiv preprint arXiv:2104.08164*, 2021. 1

[5] Bosheng Ding, Chengwei Qin, Linlin Liu, Yew Ken Chia, Shafiq Joty, Boyang Li, and Lidong Bing. Is gpt-3 a good data annotator? *arXiv preprint arXiv:2212.10450*, 2022. 1

[6] Qingxiu Dong, Lei Li, Damai Dai, Ce Zheng, Jingyuan Ma, Rui Li, Heming Xia, Jingjing Xu, Zhiyong Wu, Tianyu Liu, et al. A survey on in-context learning. *arXiv preprint arXiv:2301.00234*, 2022. 1

[7] Matthijs Douze, Alexandr Guzhva, Chengqi Deng, Jeff Johnson, Gergely Szilvasy, Pierre-Emmanuel Mazaré, Maria Lomeli, Lucas Hosseini, and Hervé Jégou. The faiss library. 2024. 4

[8] Sivan Doveh, Shaked Perek, M Jehanzeb Mirza, Wei Lin, Amit Alfassy, Assaf Arbelle, Shimon Ullman, and Leonid Karlinsky. Towards multimodal in-context learning for vision and language models. In *European Conference on Computer Vision*, pages 250–267. Springer, 2024. 3

[9] Gabriel Dulac-Arnold, Richard Evans, Hado van Hasselt, Peter Sunehag, Timothy Lillicrap, Jonathan Hunt, Timothy Mann, Theophane Weber, Thomas Degris, and Ben Coppin. Deep reinforcement learning in large discrete action spaces. *arXiv preprint arXiv:1512.07679*, 2015. 12

[10] Dyke Ferber, Georg Wölflein, Isabella C Wiest, Marta Ligero, Srividhya Sainath, Narmin Ghaffari Laleh, Omar SM El Nahhas, Gustav Müller-Franzes, Dirk Jäger, Daniel Truhn, et al. In-context learning enables multimodal large language models to classify cancer pathology images. *Nature Communications*, 15(1):10104, 2024. 2

[11] Zhaojun Guo, Jinghui Lu, Xuejing Liu, Rui Zhao, ZhenXing Qian, and Fei Tan. What makes good few-shot examples for vision-language models? *arXiv preprint arXiv:2405.13532*, 2024. 2, 3

[12] Yaru Hao, Yutao Sun, Li Dong, Zhixiong Han, Yuxian Gu, and Furu Wei. Structured prompting: Scaling in-context learning to 1,000 examples. *arXiv preprint arXiv:2212.06713*, 2022. 2

[13] Vlad Hosu, Hanhe Lin, Tamas Sziranyi, and Dietmar Saupe. Koniq-10k: An ecologically valid database for deep learning of blind image quality assessment. *IEEE Transactions on Image Processing*, 29:4041–4056, 2020. 5

[14] Hanieh Khorashadizadeh, Nandana Mihindukulasooriya, Sanju Tiwari, Jinghua Groppe, and Sven Groppe. Exploring in-context learning capabilities of foundation models for generating knowledge graphs from text. *arXiv preprint arXiv:2305.08804*, 2023. 1

[15] Hyuhng Joon Kim, Hyunsoo Cho, Junyeob Kim, Taeuk Kim, Kang Min Yoo, and Sang-goo Lee. Self-generated in-context learning: Leveraging auto-regressive language models as a demonstration generator. *arXiv preprint arXiv:2206.08082*, 2022. 2

[16] Chuanhao Li, Chenchen Jing, Zhen Li, Mingliang Zhai, Yuwei Wu, and Yunde Jia. In-context compositional generalization for large vision-language models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 17954–17966, 2024. 2

[17] Meng Li, Lin Wu, Arnold Wiliem, Kun Zhao, Teng Zhang, and Brian Lovell. Deep instance-level hard negative mining model for histopathology images. In *International conference on medical image computing and computer-assisted intervention*, pages 514–522. Springer, 2019. 2

[18] Yanshu Li. Advancing multimodal in-context learning in large vision-language models with task-aware demonstrations. In *Workshop on Reasoning and Planning for Large Language Models*. 2

[19] Lingyu Liang, Luojun Lin, Lianwen Jin, Duorui Xie, and Mengru Li. Scut-fbp5500: A diverse benchmark dataset for multi-paradigm facial beauty prediction. In *2018 24th International conference on pattern recognition (ICPR)*, pages 1598–1603. IEEE, 2018. 5

[20] Hanhe Lin, Vlad Hosu, and Dietmar Saupe. Kadid-10k: A large-scale artificially distorted iqa database. In *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–3. IEEE, 2019. 5

[21] Jiachang Liu, Dinghan Shen, Yizhe Zhang, Bill Dolan, Lawrence Carin, and Weizhu Chen. What makes good in-context examples for gpt-3? *arXiv preprint arXiv:2101.06804*, 2021. 1, 2, 3, 4, 5

[22] Yinpeng Liu, Jiawei Liu, Xiang Shi, Qikai Cheng, Yong Huang, and Wei Lu. Let's learn step by step: Enhancing in-context learning ability with curriculum learning. *arXiv preprint arXiv:2402.10738*, 2024. 1, 2, 4

[23] Katerina Margatina, Timo Schick, Nikolaos Aletras, and Jane Dwivedi-Yu. Active learning principles for in-context learning with large language models. *arXiv preprint arXiv:2305.14264*, 2023. 2

[24] Nicholas Meade, Spandana Gella, Devamanyu Hazarika, Prakhar Gupta, Di Jin, Siva Reddy, Yang Liu, and Dilek Hakkani-Tür. Using in-context learning to improve dialogue safety. *arXiv preprint arXiv:2302.00871*, 2023. 1

[25] Naila Murray, Luca Marchesotti, and Florent Perronnin. Ava: A large-scale database for aesthetic visual analysis. In *2012 IEEE conference on computer vision and pattern recognition*, pages 2408–2415. IEEE, 2012. 5

[26] Ashwinee Panda, Tong Wu, Jiachen Wang, and Prateek Mittal. Differentially private in-context learning. In *The 61st Annual Meeting Of The Association For Computational Linguistics*, 2023. 1

[27] Jakub Paplhám, Vojt Franc, et al. A call to reflect on evaluation practices for age estimation: Comparative analysis of the state-of-the-art and a unified benchmark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1196–1205, 2024. 6

[28] Kiran Purohit, V Venktesh, Sourangshu Bhattacharya, and Avishek Anand. Sample efficient demonstration selection for in-context learning. In *Forty-second International Conference on Machine Learning*. 2

[29] Chengwei Qin, Aston Zhang, Chen Chen, Anirudh Dagar, and Wenming Ye. In-context learning with iterative demonstration selection. *arXiv preprint arXiv:2310.09881*, 2023. 1, 2

[30] Ori Ram, Yoav Levine, Itay Dalmedigos, Dor Muhlgay, Amnon Shashua, Kevin Leyton-Brown, and Yoav Shoham. In-context retrieval-augmented language models. *Transactions of the Association for Computational Linguistics*, 11:1316–1331, 2023. 1

[31] Ohad Rubin, Jonathan Herzig, and Jonathan Berant. Learning to retrieve prompts for in-context learning. *arXiv preprint arXiv:2112.08633*, 2021. 2, 3

[32] Eshaan Tanwar, Subhabrata Dutta, Manish Borthakur, and Tanmoy Chakraborty. Multilingual llms are better cross-lingual in-context learners with alignment. *arXiv preprint arXiv:2305.05940*, 2023. 1

[33] Gemma Team, Aishwarya Kamath, Johan Ferret, Shreya Pathak, Nino Vieillard, Ramona Merhej, Sarah Perrin, Tatiana Matejovicova, Alexandre Ramé, Morgane Rivière, et al. Gemma 3 technical report. *arXiv preprint arXiv:2503.19786*, 2025. 5

[34] Zhengzhong Tu, Chia-Ju Chen, Li-Heng Chen, Yilin Wang, Neil Birkbeck, Balu Adsumilli, and Alan C Bovik. Regression or classification? new methods to evaluate no-reference picture and video quality models. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2085–2089. IEEE, 2021. 6

[35] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 4

[36] Qianlong Wang, Hongling Xu, Keyang Ding, Bin Liang, and Ruifeng Xu. In-context example retrieval from multi-perspectives for few-shot aspect-based sentiment analysis. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 8975–8985, 2024. 1

[37] Shuohang Wang, Yang Liu, Yichong Xu, Chenguang Zhu, and Michael Zeng. Want to reduce labeling cost? gpt-3 can help. *arXiv preprint arXiv:2108.13487*, 2021. 1

[38] Xinyi Wang, Wanrong Zhu, Michael Saxon, Mark Steyvers, and William Yang Wang. Large language models are implicitly topic models: Explaining and finding good demonstrations for in-context learning. In *Workshop on efficient systems for foundation models@ icml2023*, 2023. 1, 2, 3

[39] Xubin Wang, Jianfei Wu, Yichen Yuan, Deyu Cai, Mingzhe Li, and Weijia Jia. Demonstration selection for in-context learning via reinforcement learning. *arXiv preprint arXiv:2412.03966*, 2024. 2, 3

[40] Ziyu Wang, Tom Schaul, Matteo Hessel, Hado Hasselt, Marc Lanctot, and Nando Freitas. Dueling network architectures for deep reinforcement learning. In *International conference on machine learning*, pages 1995–2003. PMLR, 2016. 3

[41] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022. 1

[42] Wenxiao Wu, Jing-Hao Xue, Chengming Xu, Chen Liu, Xinwei Sun, Changxin Gao, Nong Sang, and Yanwei Fu. Towards reliable and holistic visual in-context learning prompt selection. *arXiv preprint arXiv:2509.25989*, 2025. 2

[43] Wenyang Xiao, Haoyu Zhao, and Lingxiao Huang. The role of diversity in in-context learning for large language models. *arXiv preprint arXiv:2505.19426*, 2025. 2

[44] Hongling Xu, Qianlong Wang, Yice Zhang, Min Yang, Xi Zeng, Bing Qin, and Ruifeng Xu. Improving in-context learning with prediction feedback for sentiment analysis. *arXiv preprint arXiv:2406.02911*, 2024. 1

[45] Lu Xu, Jinhai Xiang, and Xiaohui Yuan. Transferring rich deep features for facial beauty prediction. *arXiv preprint arXiv:1803.07253*, 2018. 6

[46] Jinghan Yang, Shuming Ma, and Furu Wei. Auto-icl: In-context learning without human supervision. *arXiv preprint arXiv:2311.09263*, 2023. 2

[47] Li Yang, Zengzhi Wang, Ziyan Li, Jin-Cheon Na, and Jianfei Yu. An empirical study of multimodal entity-based sentiment analysis with chatgpt: Improving in-context learning via entity-aware contrastive learning. *Information Processing & Management*, 61(4):103724, 2024. 1

[48] Zhao Yang, Yuanzhe Zhang, Dianbo Sui, Cao Liu, Jun Zhao, and Kang Liu. Representative demonstration selection for in-context learning with two-stage determinantal point process. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 5443–5456, 2023. 2

[49] Liang Yao. Large language models are contrastive reasoners. *arXiv preprint arXiv:2403.08211*, 2024. 3

[50] Jiacheng Ye, Zhiyong Wu, Jiangtao Feng, Tao Yu, and Lingpeng Kong. Compositional exemplars for in-context learning. In *International Conference on Machine Learning*, pages 39818–39833. PMLR, 2023. 2

[51] Xiaohua Zhai, Basil Mustafa, Alexander Kolesnikov, and Lucas Beyer. Sigmoid loss for language image pre-training. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 11975–11986, 2023. 3

[52] Wenxuan Zhang, Yue Deng, Bing Liu, Sinno Jialin Pan, and Lidong Bing. Sentiment analysis in the era of large language models: A reality check. *arXiv preprint arXiv:2305.15005*, 2023. 1

[53] Yiming Zhang, Shi Feng, and Chenhao Tan. Active example selection for in-context learning. *arXiv preprint arXiv:2211.04486*, 2022. 2, 3

[54] Yuanhan Zhang, Kaiyang Zhou, and Ziwei Liu. What makes good examples for visual in-context learning? *Advances in Neural Information Processing Systems*, 36:17773–17794, 2023. 2, 3

[55] Zhifei Zhang, Yang Song, and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017. 5

[56] Zheng Zhang, Shaocheng Lan, Lei Song, Jiang Bian, Yexin Li, and Kan Ren. Learning to select in-context demonstration preferred by large language model. *arXiv preprint arXiv:2505.19966*, 2025. 2

[57] Yucheng Zhou, Xiang Li, Qianning Wang, and Jianbing Shen. Visual in-context learning for large vision-language models. *arXiv preprint arXiv:2402.11574*, 2024. 2

[58] Yan Zhu, Huan Ma, and Changqing Zhang. Exploring task-level optimal prompts for visual in-context learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 11031–11039, 2025. 1