

# Improving Image-recognition Edge Caches with a Generative Adversarial Network

Guilherme B. Souza\*, Roberto G. Pacheco\*, Rodrigo S. Couto\*

\*Universidade Federal do Rio de Janeiro, GTA/PEE-COPPE/DEL-Poli, Rio de Janeiro, RJ, Brazil

Email: bergman@gta.ufrj.br, pacheco@gta.ufrj.br, rodrigo@gta.ufrj.br

**Abstract**—Image recognition is an essential task in several mobile applications. For instance, a smartphone can process a landmark photo to gather more information about its location. If the device does not have enough computational resources available, it offloads the processing task to a cloud infrastructure. Although this approach solves resource shortages, it introduces a communication delay. Image-recognition caches on the Internet’s edge can mitigate this problem. These caches run on servers close to mobile devices and stores information about previously recognized images. If the server receives a request with a photo stored in its cache, it replies to the device, avoiding cloud offloading. The main challenge for this cache is to verify if the received image matches a stored one. Furthermore, for outdoor photos, it is difficult to compare them if one was taken in the daytime and the other at nighttime. In that case, the cache might wrongly infer that they refer to different places, offloading the processing to the cloud. This work shows that a well-known generative adversarial network, called ToDayGAN, can solve this problem by generating daytime images using nighttime ones. We can thus use this translation to populate a cache with synthetic photos that can help image matching. We show that our solution reduces cloud offloading and, therefore, the application’s latency.

1

## I. INTRODUCTION

Mobile devices commonly use cameras to extract useful information and provide users with a better understanding of their surroundings. For example, smartphone applications, such as *Google Lens*<sup>2</sup>, identify objects or landmarks targeted by the device’s camera. Once the target is recognized, *Google Lens* provides detailed information about it. In the case of landmarks, the application may return the location name and some information about the place (e.g., historical data). Another example is the way cameras are used in intelligent vehicles, in which decision-making processes take visual data into account [1], [2]. Image recognition requires high computational and energy utilization. These requirements are often incompatible with mobile devices [3]. Hence, a solution is to offload image recognition to a cloud server, which does not have the same limitations. The server, in turn, processes the image and returns the result to the device.

Although sending the image to a cloud server solves resource constraint problems, it introduces communication

delay. This delay increases the application’s latency and can be prohibitive to real-time scenarios [1]. Edge computing aims to reduce this problem and other ones caused by the cloud [4]. With this paradigm, we can place computational resources on the edge of the Internet (e.g., WiFi access points and base stations). It is thus possible to reduce the communication delay and the amount of data sent to the cloud.

An edge server can deploy image-recognition (IR) caches to reduce applications’ latency and network traffic. Cachier [5] and FoggyCache [6] both propose IR caches on the edge. Their systems intercept requests from devices before offloading to the cloud. A device can request, for example, information about a photographed location or an object recognition in an image. When an image arrives at the edge, it is compared with other images stored in the cache. The system checks if the received image is similar to any of the stored ones. If so, it returns to the user the result stored for the matched image. Reusing results in cache avoids offloading and, therefore, reduces the application’s latency. If the system cannot find any similar match, the request is forwarded to the cloud. Once the cloud solves the request, the result is returned to the device and may be stored in cache for future use.

IR caches are based on the idea that different users can send a request with similar images. Various methods can assess the similarity between images. FoggyCache and Cachier use feature detection algorithms such as ORB (Oriented FAST and Rotated BRIEF) [7] and SIFT (Scale-invariant feature Transform) [8]. More recent approaches opt for machine-learning techniques to encode images [9]. The encodings can be later compared to obtain a similarity value between images.

Both feature detection algorithms and machine learning approaches can have poor performance when comparing images of the same place or object but under different illumination conditions [10]. It means that an image cache may fail to reuse results from similar requests. For example, a cache that contains stored nighttime images will not perform well when receiving daytime requests. Consequently, the edge will unnecessarily offload to the cloud, even when the cache contains information about the user’s request.

This paper<sup>3</sup> proposes a solution to make a cache system more robust to illumination variations. Specifically, we

<sup>1</sup>©2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

<sup>2</sup><https://lens.google.com>

<sup>3</sup>Some fragments and ideas of this work are based on our preliminary paper (<https://sol.sbc.org.br/index.php/courb/article/view/17116>), published in Portuguese in a Brazilian workshop. This utilization is permitted by the Brazilian publisher, as seen in <https://sol.sbc.org.br/index.php/indice/conduca>.

consider an urban scenario in which the cache contains nighttime images and receives daytime requests. We achieve robustness through an image translation technique, called ToDayGAN [11], which transforms real nighttime images into synthetic day images using Generative Adversarial Networks (GANs) [12]. The synthetic images are stored in the cache, increasing the odds of solving the requests on the edge server.

The experiments presented in this paper use a common dataset in the literature that contains vehicle-captured images [13]. We evaluate two different scenarios of an IR cache that receives daytime requests. One has only real nighttime images stored. The other one contains synthetic daytime images generated from these nighttime images. Results show that the storage of synthetic images reduces cloud offloading, and therefore requests’ latency. Thus, our proposal can improve the latency of applications dependent on IR edge caches.

This paper is organized as follows. Section II describes related work. Section III details our approach. Section IV describes the evaluation methodology, while Section V present our results. Finally, Section VI concludes the work.

## II. RELATED WORK

An IR cache exploit requests’ locality in mobile applications. This locality can be observed, for example, in image recognition applications, in which geographically close users often request content related to the same object or landmark [6]. Cachier [5] is an IR cache that uses ORB to detect local features from the images included in the requests. These features are then used to train a classification model. Hence, new requests can be classified as possible objects stored in the cache that reuses the stored result in the response. The reuse made possible by Cachier achieves an average latency of up to three times less than the exclusive use of the cloud. Precog [14] extends Cachier’s proposal, spreading parts of the cache on the mobile devices and the edge. Precog can thus reduce the application’s latency up to five times less than an edge-only cache. This is true since Precog can efficiently reuse the content cached on mobile devices, avoiding unnecessary requests to external edge infrastructure.

FoggyCache [6] proposes a cache system that can reuse computation for several deep neural networks (DNNs) models. To this end, FoggyCache encodes the inputs as high dimensional vectors stored in the cache. For IR applications, it uses SIFT to detect and extract features from images that are later combined to form encoding vectors. FoggyCache also develops a method to compare the cached vectors with the new inputs to identify similar requests accurately. As a result, FoggyCache reduces an application latency by a factor of 3 to 10.

Shadow Puppets [9] uses a neural network as a feature extractor instead of the traditional methods, such as ORB and SIFT. The results of Shadow Puppets show that the neural network can efficiently identify similarities between cached data and the new ones. Thus, it shows that using a neural network as a feature extractor can increase the cache’s reuse.

Inspired by Shadow Puppets, our work also uses a neural network-based approach for image retrieval in the cache.

Prior works optimize the image retrieval process or cache replacement policies, reducing the required search time and cloud offloading. In contrast, our work tackles the problem of the reduced performance of an IR cache when receiving queries under very different lighting conditions. Our proposal of storing synthetic images using a ToDayGAN is thus agnostic to IR cache solutions used in the previous works.

## III. IMAGE-RECOGNITION EDGE CACHE

This section presents our approach by describing the cache system and detailing the procedure of generating images.

### A. System Overview

An IR cache works similarly to a web cache, such as an HTTP proxy [5]. A web cache stores the contents of web pages recently requested by users. Each content is indexed with an URL. The proxy looks for the request’s URL to determine if the required content is stored in the cache. If so, the web cache returns the content to the user, avoiding requests to an Internet server. Otherwise, the web cache requests the desired content to the Internet server. However, IR caches do not perform an exact search with an identifier such as an URL. For example, in a geolocation application, a user sends an image of a specific location and waits for the server or cache to return information about the requested location. In this case, even if multiple users photograph the exact location, these images will have differences, such as lighting conditions or the angle at which the image is captured. The search in IR caches evaluates the similarities between requested images and the cached images extracting their descriptors using neural networks or traditional approaches, such as SIFT. These descriptors allow the cache to compute a similarity metric between the requested image and the stored ones. The reuse is then accomplished by finding similar images stored in the cache and returning the stored content.

We apply the NetVLAD [15] neural network to encode the input image, extracting its descriptors. NetVLAD outputs a 16k dimension encoding vector for every input image. We thus compute the similarity between the encoding vector of the input image and the cached ones using Euclidean distance as the similarity metric. Rather than storing images themselves, the cache stores their encoding and their associated content.

Figure 1 illustrates our system. The edge receives an image sent from a mobile device and extracts its encoding using NetVLAD. This encoding is thus used as input to a  $k$ -nearest-neighbors search [16]. This search finds the  $k$  most similar encoding stored in the cache. Before returning the content of the most similar image, the system performs a threshold comparison to assess the correctness of the result. This comparison is based on Lowe’s Ratio Test [17]. To this end, we obtain the distances  $d_1$  and  $d_2$ , the euclidean distances between the requested image’s encoding and two close ones. In Figure 1, we show an example with the five ( $k = 5$ ) closest encodings. The closest encoding is  $d_1$ . In our example, the

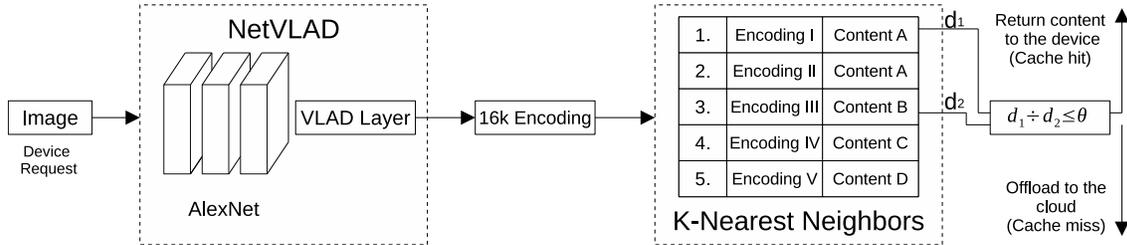


Fig. 1. Our IR cache system overview. First, NetVLAD generates an encoding for the request image. This encoding is then compared with the cached ones to find the most likely requested content. The edge returns the content to the device (cache hit) or forward the request to the cloud (cache miss).

two closest encodings have the same content  $A$ . Distance  $d_2$  is hence obtained from the third closest encoding. Afterwards, we compute the ratio  $r = \frac{d_1}{d_2}$  and compare it with a predefined threshold  $\theta$ . If  $r \leq \theta$ , the search result is considered incorrect, so the IR cache offloads the image to the cloud (cache miss). Otherwise, the IR cache returns the content associated with the most similar image in the cache (cache hit). We employ this system in a geolocation application. In this case, the contents are strings indicating the requested images' location.

The cloud uses a mechanism similar to the cache to reply to requests. However, the cloud has more images than the edge and may use more sophisticated image retrieval approaches. As our focus is only the IR cache, we emulate cloud image retrieval as follows. First, the cloud contains both daytime and nighttime images. Furthermore, the cloud has the corresponding encoding for the same photos used in the requests. It can be an unpractical situation but helps emulate a better image retrieval performed in the cloud. The images on the edge cache are different from the ones on the requests. Thus, the cloud can achieve much higher precision than the cache. In real scenarios, the cloud may not have access to the complete dataset beforehand and, therefore, will resort to another approach to achieve high precision. In [6], for example, the cloud runs an accurate deep neural network.

### B. Image retrieval

Image retrieval can use manually crafted feature extractors. These methods work by locating key points in the image and generating vectors (descriptors) to describe them. The similarity is hence determined by computing a distance between images' descriptors, such as Euclidean distance.

Similar to Shadow Puppets, we use a neural network as an alternative to manually crafted feature extractors. Instead of extracting multiple features from an image, we use NetVLAD [15] to generate a single encoding descriptor for the whole image. Analogous to ORB and SIFT descriptors, the Euclidean distance between the encodings is used to determine the similarity metric between images.

NetVLAD combines both hand-engineered and machine learning approaches to propose an architecture for accurate place recognition. It consists of a convolutional neural network (CNN) connected at its end to a VLAD (Vector of Locally Aggregated Descriptors) layer, inspired by the VLAD

descriptor aggregator commonly used in the image retrieval literature. They train the CNN to generate descriptors with small Euclidean distances for geographically close images and vice versa. This training allows easy comparison of NetVLAD descriptors in image retrieval. The VLAD layer of NetVLAD can be easily plugged into any CNN to form different versions of the architecture. In this work, we use the AlexNet version of NetVLAD, although a VGG-16 can achieve better results [15]. We choose the AlexNet as a pessimistic approach since it can better fit hardware-constrained edge caches with low latency requirements [18]. Our training follows the same methodology of [15], using the Pittsburgh (Pitts250k) [19] dataset.

NetVLAD has significant improvements in image retrieval in diverse databases compared to baseline models [15]. Still, Anoosheh *et al.* [11] and our paper show that image translation techniques improve, even more, the performance of image retrieval by storing synthetic daytime images in the database.

### C. Generation of synthetic images

Generative Adversarial Networks (GANs) [12] is an unsupervised learning algorithm designed to solve generative modeling problems. In generative modeling, the training samples given by  $\mathbf{x}$  are drawn from an unknown distribution  $p_{\text{data}}(\mathbf{x})$ . The goal of a generative modeling algorithm is to learn a probability distribution  $p_{\text{model}}(\mathbf{x})$  that approximates  $p_{\text{data}}(\mathbf{x})$ . GANs have recently achieved considerable advances in computer vision applications, such as image restoration [20], synthetic image generation [21], [22], and image-to-image translation [23]. The latter consists of learning data distributions to convert from a source domain  $X$  to a target domain  $Y$ . For example, a CycleGAN [24] model learns a painter's style, using a collection of their pictures, and can convert an ordinary photo to the learned style.

Our work employs a ToDayGAN [11] to show that GANs can improve IR caches. ToDayGAN is an image-to-image translation model, performing the translation between daytime and nighttime image domains. This model, which has a structure based on a CycleGAN, is trained on an urban images dataset gathered by an autonomous vehicle day and night.

ToDayGAN design aims to aid geolocation image retrieval [11]. ToDayGAN's proposal considers that the reference database stores daytime images and that the system receives nighttime photos. Thus, before searching for the

most similar image into the database, ToDayGAN transforms nighttime images into synthetic daytime images to improve the performance. Next, ToDayGAN extracts features of the transformed image to compare with the stored ones and finds the best match. Our work applies ToDayGAN in IR caches also with geolocation application in mind. We exploit the ToDayGAN’s capability of matching images under different illumination conditions to reduce cloud offloading and the latency in cache-dependent applications. Our work follows the same idea used in ToDayGAN [11], considering turning daytime images into nighttime ones. Moreover, we translate the images already stored in the IR cache instead of translating the photos in real-time. This translation should occur when the cache server is idle to not affect the system’s performance.

#### IV. METHODOLOGY

This work uses the Oxford RobotCar dataset [13], composed of frames from several videos recorded on the same route in Oxford, England. An autonomous vehicle gathers these frames through cameras attached to its front, back, and sides. This vehicle repeats the same path for several days, recording different weather and lighting conditions. The dataset is divided into days of recording. We then sample images from one recording with overcast weather and use them as daytime requests. Images from two different night recordings are sampled to form the cached content. We use the ToDayGAN model trained in [11] for the image-to-image translation. It also uses the Oxford RobotCar, employing part of this dataset for its training. Although our experiments use the same dataset, the images used to train this model are not part of the test set for evaluating our proposal.

Each image of Oxford RobotCar has GPS data indicating its coordinates. Therefore, we group the dataset images by location to consider a geolocation application. The vehicle’s route is split into segments of five meters each. The experiments assume that images that originate from the same segment belong to the same place and hence should receive the same result from the application when sent as a request.

Our experiments consider two main scenarios. In the first one, referred to as “Night”, the cache stores only nighttime images. The other one, referred to as “GAN”, evaluates the performance of the same cache system but instead stores only synthetic daytime images generated by the ToDayGAN. These synthetic images are generated through the translation of the nighttime photos used in the first scenario. In both cases, the cache only receives real daytime images as requests.

We evaluate three metrics. The precision is the percentage of requests that the cache or the cloud have correctly identified their images’ location. The recall is the percentage of correctly identified requests whose content is cached. Finally, we evaluate the application’s latency, described later in Section V-C.

#### V. RESULTS

This section evaluates our proposal. We use the experiments of Section V-A to choose an adequate threshold  $\theta$  value. Afterward, we analyze in Section V-B the impact of the

number of cached places in precision and recall. Finally, Section V-C evaluates the benefits of our proposal to a geolocation application’s latency.

##### A. Threshold choice

This first experiment evaluates the IR cache for different thresholds, using the “Night” and “GAN” scenarios and the case where no cache is available. For each threshold, we perform ten experimental rounds. In each round, we choose 100 possible places from the dataset. We then choose to cache 50% of these places. This percentage allows verifying if the chosen thresholds correctly discard results of images belonging to places not cached. For each cached place, we have eight stored encodings corresponding to eight different images. We choose this value arbitrarily to show that our solution is still necessary even when the cache stores many images. In this first experiment, we use 150 images requests, chosen randomly from the 100 possible locations of each round. All results of this work are average values of ten rounds with confidence intervals of 95%.

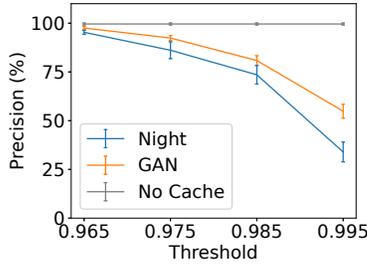
Figure 2(a) shows the precision results that indicate the percentage of requests in which the system retrieves the correct place. An incorrect image retrieval occurs when the cache returns the information of a location different from the requested one. This false retrieval also occurs when the edge offloads a request to the cloud, and the cloud returns the wrong place. This last case is rare since the cloud performs an almost ideal image retrieval, as described in Section III. To illustrate this behavior, the “No Cache” curve of Figure 2(a) shows the precision when the edge offloads all requests to the cloud, and it is not dependent on the threshold.

For the “Night” and “GAN” scenarios, Figure 2(a) shows that the precision shifts down as we increase the threshold value. The recall curve of Figure 2(a) represents the percentage of requests that the cache returns the correct content, considering only those whose corresponding place is cached [5]. Note that Figures 2(a) and 2(a) have an opposite behavior regarding the threshold. It happens because a small threshold results in a small number of erroneous results being returned by the cache and hence a high precision. At the same time, it also forces unnecessary offloading of correct results to the cloud, lowering the recall. Changing this value thus allows fine-tuning the trade-off between recall and precision.

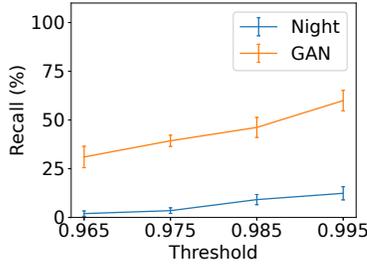
To achieve a high recall while maintaining a high precision (around 90%), we set the threshold value to 0.975 for the remaining experiments. It is also important to note that, although the “Night” scenario reaches a precision close to the “GAN”, this last one achieves a higher recall. It is the first evidence that the image translation technique can increase an IR cache performance by reducing the amount of offloaded requests to the cloud. The following experiments confirm the technique’s efficiency for a different number of cached places.

##### B. Cache coverage variation

The following experiment uses a similar methodology to the first one. This time, we fix the threshold value at 0.975



(a) Precision.



(b) Recall

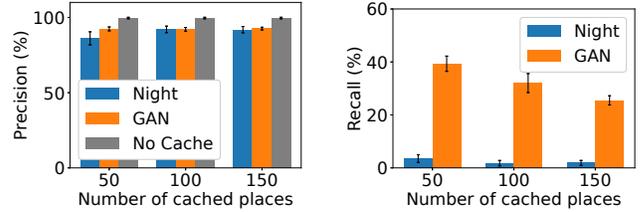
Fig. 2. Threshold variation with 50% of cached places.

and vary the number of places stored in the cache. This number can be seen as cache coverage. A cache that stores images from many locations can receive requests from a larger geographical area than a cache that stores fewer places. Depending on its configuration, a cache may cover nearby streets, neighborhoods, cities, etc. As in Section V-A, the cache has 50% of all possible locations with eight images for each. Hence, for every number of cached places, we set the number of possible locations accordingly. As in Section V-A, the total number of requests is three times the number of cached locations. Note that for 50 cached places, we have the same experiment of Section V-A when  $\theta = 0.975$ , with 100 possible locations and 150 requests.

Figure 3(b) shows that the recall decays as we increase the number of cached places. We expect this behavior because the cache needs to distinguish an image between more possibilities when more locations are cached. On the other hand, Figure 3(a) shows that altering the number of cached places causes no significant impact on the precision. Finally, as in the first experiment, the cache with synthetic images achieves a higher recall for any configuration. We fix the number of cached places in 50 for the subsequent evaluation to represent a small edge cache that receives only nearby requests.

### C. Latency evaluation

The final experiment evaluates the impact of the recall improvements on the latency. We use the same methodology as before, but now we measure the latency. We define this metric as the total time elapsed since the arrival of a request on the edge until the obtention of its response. We disregard the communication delay between the device and the edge



(a) Precision.

(b) Recall.

Fig. 3. Cache coverage variation with 50% of cached places and  $\theta = 0.975$ .

because it is agnostic to the cache solution.

We set up the experiment with two machines working as the edge cache and the cloud. The cache runs a Debian 9.13 operating system. It has an Intel i5-9600K CPU with six cores at 3.70 GHz and an NVIDIA GeForce RTX 2080 Ti GPU. The cloud server is an Amazon AWS EC2 `g4dn.xlarge` instance, with a Deep Learning AMI (Ubuntu 18.04) 51.0, four vCPUs from an Intel Cascade Lake CPU, and an NVIDIA Tesla T4 GPU. The edge runs in our laboratory in Rio de Janeiro, Brazil. We instantiate the cloud machine in two different AWS regions to consider different network conditions. One region is far from the edge, located in Ohio, USA. The other one is in São Paulo, Brazil, the closest AWS region to Rio de Janeiro at the date of this paper’s writing.

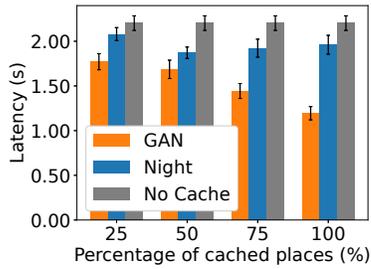
Table I shows the network conditions between the cache and the cloud for each region. To this end, we apply `iPerf` and `ping` tools to measure the average RTT (Round-Trip Time) and throughput. This table only illustrates the network conditions. Their values depend on the Internet performance, suffering fluctuation during the day.

TABLE I  
NETWORK CONDITIONS FOR EACH CLOUD REGION.

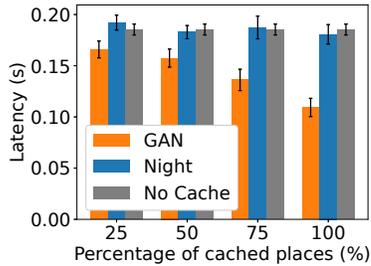
AWS Region	Location	Throughput	RTT
us-east-2	Ohio	86.9 Mbit/sec	145.65 ms
sa-east-1	São Paulo	93.4 Mbits/sec	12.39 ms

Figure 4 presents the latency results for the two AWS regions. In the x-axis, we vary the percentage of cached places. The previous experiments always fixed this percentage in 50%. We now set the number of cached places to 50 and change the number of possible locations to induce each percentage. This value is varied since it has a direct impact on latency. If the cache has a higher fraction of the total locations, it will result in less offloading to the cloud and, therefore, lower latency. This percentage changes over time in a production system and depends on the cache’s replacement policy, which is orthogonal to our proposal. The results show that storing synthetic daytime images reduces the application’s latency even for very different network conditions. This reduction is more expressive as we increase the percentage of requests belonging to cached places. Therefore fewer requests are offloaded to the cloud. Figure 4(b) shows the results for the region with the lowest offloading penalty. Even in this case,

the GAN reduces the latency from 13.5% to 40.6% compared to using real nighttime images.



(a) Ohio (USA) - us-east-2.



(b) São Paulo (Brazil) - sa-east-1

Fig. 4. Latency evaluation for  $\theta = 0.975$  and 50 cached places.

## VI. CONCLUSIONS

This paper has shown that an image retrieval (IR) cache containing nighttime images exhibits low performance when receiving daytime images as requests. We solve this problem using ToDayGAN [11], a well-known GAN-based image generator, to translate the IR cache’s content into daytime images. With this approach, the cache can reuse more computation, avoiding cloud offloading. Our experiments have used an actual cloud provider to show that the proposal reduces the application’s latency. This effect is thus our main benefit, being particularly useful to latency-sensitive applications as they can experience a drop in user experience. These applications may even be rendered unfeasible by high response times. For future works, we plan on evaluating our proposal considering other concerns typical of cache literature, such as replacement policy. Furthermore, our solution is independent of the cache implementation. Hence, we may also investigate the image translation technique efficiency for different datasets and image retrieval mechanisms.

## ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. It was also supported by CNPq, FAPERJ Grants E-26/203.211/2017, E-26/211.144/2019, and E-26/201.300/2021, and FAPESP Grant 15/24494-8.

## REFERENCES

- [1] M. G. Bechtel, E. McEllhiney, M. Kim, and H. Yun, “DeepPicar: A low-cost deep neural network-based autonomous car,” in *IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA)*, 2018, pp. 11–21.
- [2] H. Kim, Y. Lee, B. Yim, E. Park, and H. Kim, “On-road object detection using deep neural network,” in *IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, 2016, pp. 1–4.
- [3] E. Cuervo, A. Balasubramanian, D.-k. Cho, A. Wolman, S. Saroiu, R. Chandra, and P. Bahl, “Maui: making smartphones last longer with code offload,” in *International Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2010, pp. 49–62.
- [4] R. G. Pacheco, R. S. Couto, and O. Simeone, “Calibration-aided edge inference offloading via adaptive model partitioning of deep neural networks,” in *IEEE International Conference on Communications (ICC)*, 2021, pp. 1–6.
- [5] U. Drolia, K. Guo, J. Tan, R. Gandhi, and P. Narasimhan, “Cachier: Edge-caching for recognition applications,” in *IEEE International Conference on Distributed Computing Systems (ICDCS)*, 2017, pp. 276–286.
- [6] P. Guo, B. Hu, R. Li, and W. Hu, “FoggyCache: Cross-device approximate computation reuse,” in *International Conference on Mobile Computing and Networking (MobiCom)*, 2018, pp. 19–34.
- [7] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: An efficient alternative to SIFT or SURF,” in *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2011, pp. 2564–2571.
- [8] D. G. Lowe, “Object recognition from local scale-invariant features,” in *IEEE/CVF International Conference on Computer Vision (ICCV)*, 1999, pp. 1150–1157.
- [9] S. Venugopal, M. Gazzetti, Y. Gkoufas, and K. Katrinis, “Shadow Puppets: Cloud-level accurate AI inference at the speed and economy of edge,” in *USENIX Workshop on Hot Topics in Edge Computing (HotEdge 18)*, 2018.
- [10] T. Sattler, W. Maddern, C. Toft, A. Torii, L. Hammarstrand, E. Stenborg, D. Safari, M. Okutomi, M. Pollefeys, J. Sivic *et al.*, “Benchmarking 6dof outdoor visual localization in changing conditions,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 8601–8610.
- [11] A. Anosheh, T. Sattler, R. Timofte, M. Pollefeys, and L. Van Gool, “Night-to-day image translation for retrieval-based localization,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2019, pp. 5958–5964.
- [12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” *Advances in neural information processing systems*, vol. 27, 2014.
- [13] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, “1 year, 1000 km: The oxford robotcar dataset,” *The International Journal of Robotics Research*, vol. 36, no. 1, pp. 3–15, 2017.
- [14] U. Drolia, K. Guo, and P. Narasimhan, “Precog: Prefetching for image recognition applications at the edge,” in *ACM/IEEE Symposium on Edge Computing (SEC)*, 2017, pp. 1–13.
- [15] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, “Netvlad: Cnn architecture for weakly supervised place recognition,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 5297–5307.
- [16] J. Johnson, M. Douze, and H. Jégou, “Billion-scale similarity search with gpus,” in *IEEE Transactions on Big Data*, 2019, pp. 535–547.
- [17] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” vol. 60, no. 2, 2004, pp. 91–110.
- [18] S. Bianco, R. Cadene, L. Celona, and P. Napolitano, “Benchmark analysis of representative deep neural network architectures,” *IEEE Access*, vol. 6, pp. 64 270–64 277, 2018.
- [19] A. Torii, J. Sivic, T. Pajdla, and M. Okutomi, “Visual place recognition with repetitive structures,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 883–890.
- [20] R. A. Yeh, C. Chen, T. Yian Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, “Semantic image inpainting with deep generative models,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5485–5493.
- [21] A. Elgammal, B. Liu, M. Elhoseiny, and M. Mazzone, “Can: Creative adversarial networks, generating” art” by learning about styles and deviating from style norms,” *arXiv preprint arXiv:1706.07068*, 2017.

- [22] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *International Conference on Learning Representations*, 2018.
- [23] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1125–1134.
- [24] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2017, pp. 2223–2232.