A SUPERPOINT NEURAL NETWORK IMPLEMENTATION FOR ACCURATE FEATURE EXTRACTION IN UNSTRUCTURED ENVIRONMENTS

Georgios Petrakis
Technical University of Crete, University Campus, Akrotiri, 73100 Chania, Greece
email: gpetrakis2@tuc.gr

KEY WORDS: Feature extraction, deep neural networks, unstructured environments

ABSTRACT:

Feature extraction plays a crucial role in visual localization, SLAM (Simultaneous Localization and Mapping) and autonomous navigation, by enabling the extraction and tracking of distinctive visual features for both mapping and localization tasks. However, most of the studies, investigate the efficiency and performance of the algorithms in urban, vegetated or indoor environments and not in unstructured environments which suffers by poor information in visual cues where a feature extraction algorithm or architecture could base on. In this study, an investigation of SuperPoint architecture's efficiency in keypoint detection and description applied to unstructured and planetary scenes was conducted, producing three different models: (a) an original SuperPoint model trained from scratch, (b) an original fine-tuned SuperPoint model, (c) an optimized SuperPoint model, trained from scratch with the same parametarization as the corresponding original model. For the training process, a dataset of 48 000 images was utilized representing unstructured scenes from Earth, Moon and Mars while a benchmark dataset was used aiming to evaluate the model in illumination and viewpoint changes. The experimentation proved that the optimized SuperPoint model provides superior performance using repeatability and homography estimation metrics, compared with the original SuperPoint models, and handcrafted keypoint detectors and descriptors.

1. INTRODUCTION

Feature extraction plays a crucial role in visual localization, SLAM (Simultaneous Localization and Mapping) and autonomous navigation, by enabling the extraction and tracking of distinctive visual features for both mapping and localization tasks (Nixon & Aguado 2019). These features represent distinctive points or structures in the environment including corners or keypoints, which provide a compact and efficient representation of the environment, while serve as reference points for associating sensor data with the map and estimating the robot's pose. Furthermore, feature extraction helps to handle challenges such as occlusions, dynamic environments, or changing lighting conditions, by focusing on stable and discriminative visual cues.

Several handcrafted algorithms including keypoint detectors and descriptors such as Shi-Tomasi (Shi & Tomasi 1993), Harris (Harris & Stephens 1988), Fast (Rosten & Drummond 2006), Orb (Rublee, 2011), Sift (Lowe, 2004) etc and deep learning architectures such as SuperPoint (DeTone, 2018), UnsuperPoint (Christiansen, 2019), Lf-net (Ono, 2018), D2-net (Dusmanu, 2019), etc have been proposed in the literature (Bojanić, 2019). Handcrafted keypoint detectors rely on designed filters or mathematical operations that are based on gradient-based or intensity-based techniques, while attempt to maintain their reliability in scale, rotation, and viewpoint changes (Isık & Ozkan 2015). On the other hand, deep learning architectures, create response maps aiming to detect interest points while subsequently learn representations of each keypoint using either local patches centred on the predicted keypoints or the entire image utilizing the pixel-level keypoint locations (Ma, 2021).

However, most of the studies, investigate the efficiency and performance of the algorithms in urban, vegetated or indoor environments and not in unstructured environments which suffers by poor information in visual cues where a feature extraction algorithm or architecture could base on (Guastella & Muscato 2021).

Moreover, although there is a plethora of datasets for evaluation and training of feature extraction algorithms such as HPatches (Balntas, 2017), Aachen (Sattler, 2008), COCO (Lin, 2014), Google Landmarks (Noh, 2017), etc, they focus on urban, indoor or vegetated environments while the datasets which represent unstructured scenes are quite few and they are designed mainly for SLAM (Simultaneous Localization and Mapping) evaluation (Meyer, 2021, Furgale, 2012, Giubilato, 2022, Hewitt, 2018) and not for training or keypoint detector or descriptor evaluation.

In this study, a SuperPoint neural network (DeTone, 2018) is implemented, optimized and trained in order to accurately conduct feature extraction in unstructured environments, focused on rocky and sandy scenes. For the training process, a dataset of 48 000 images was utilized (Petrakis & Partsinevelos 2023) representing unstructured and planetary scenes from Earth, Moon and Mars. Concerning images from Earth, were captured from construction sites, mountainous areas, sandy beaches and a quarry, while the images from Mars were collected by a publicly available dataset of NASA which includes rover-based images, captured by Mars Science Laboratory (MSL, Curiosity) rover. Regarding the Moon, the dataset includes artificial images, created by Keio University in

https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-1215-2023 |

Japan. Concerning the learning procedure, the MagicPoint detector, the standalone part of SuperPoint detector, was trained in three phases, one time with synthetic data and two times using the aforementioned dataset using homographic adaptation, a technique to increase the efficiency of the architecture in geometric transformations. Finally, the SuperPoint neural network was trained based on the weights of MagicPoint in order to fine-tune the keypoint detector and train the descriptor. Three different models were produced using the aforementioned dataset: (a) an original SuperPoint model, trained from scratch, (b) an original fine-tuned SuperPoint model, (c) an optimized model, trained from scratch. The models were evaluated using a benchmark dataset (Petrakis & Partsinevelos, 2023), designed for unstructured environments including earthy and planetary scenes, aiming to test the accuracy in illumination and viewpoint changes. The experimentation proves that the optimized SuperPoint model provides satisfactory results in keypoint detection and description, compared with the original SuperPoint and popular handcrafted detectors and descriptors.

2. MATERIALS AND METHODS

The main goal of this study, is the implementation of a feature extraction architecture, capable to detect and describe keypoints in challenging unstructured environments or planetary scenes with lack of visual cues and intense illumination changes. To deal with these challenges, SuperPoint (DeTone et al. 2018), a state-of-the-art neural network which outperforms handcrafted and deep learning feature extractors (Bojanic, 2019, Liu, 2022) was implemented and improved.

2.1 SuperPoint architecture

Superpoint is a fully convolutional neural network, composed by an encoder-decoder architecture which is performed using full-sized images as input. At first, a shared encoder, based on VGG neural network (Simonyan & Zisserman 2015) is utilized aiming to reduce the image dimensionality using three max-polling operations, extracting image cells in a size of $H_c = H / 8$ and $W_c = W / 8$ where H and W are the height and width of an image. The extracted tensor is imported in two decoders, one of which acts as a keypoint detector and the other one as a descriptor (fig 1).

Concerning the keypoint detector decoder, it undertakes the reconstruction of the full-sized image, extracting the probability of a keypoint existence in each pixel. Initially, it forms a tensor $X \in \mathbb{R}^{H_c W_c \times 65}$ where 65 channels are composed by 64 non-overlapping 8x8 pixel cells and an extra cell, called “no interest point dustbin” (DeTone, 2018). Subsequently, this tensor is imported to a “softmax function” where the dustbin cell is removed while the resulted tensor is reshaped to a full-sized image output ($\mathbb{R}^{H x W}$) after a “reshape operation” (fig. 1). It’s worth noting that the detector decoder doesn’t upsample the full resolution of the image using transposed convolution techniques such as Unet (Ronneberger, 2015) due to high demands on computing resources while according to DeTone et al. 2018, these upsampling techniques are able to introduce checkerboard artifacts. Instead a”sub-pixel convolution” (Shi, 2016) is utilized, which doesn’t include training parameters, aiming to reduce the computation process.

Regarding the descriptor decoder, it computes a tensor $\mathbb{R}^{H_c W_c \times D}$ where D is the descriptor length equal to 256 while via two convolutional layers, it extracts fixed feature maps in a shape of $I_{desc}{}^{H_c W_c \times D}$. The feature maps are reconstructed to the full-sized dimensions through a bi-linear interpolation while afterwards, the L2 norm operation is performed aiming to calculate the unit length of the descriptors (fig. 1). It’s worth noting that the original SuperPoint architecture utilizes bi-cubic interpolation instead of bi-linear. However, in case of unstructured environments, it was observed that bi-linear interpolation provided similar accuracy while reducing the computation process compared with bi-cubic interpolation.

SuperPoint utilizes a unified loss function which is composed by the loss function of keypoint detector ($\mathcal{L}_p$) and the loss function of the descriptor ($\mathcal{L}_d$). SuperPoint uses pairs of wrapped images with the predicted keypoint locations and the corresponding transformation matrices or homography, utilized as ground truth. The unified loss function is presented in equation (1):

$$\mathcal{L}(X, X', D, D'; Y, Y', S) = \mathcal{L}_p(X, Y) + \mathcal{L}_p(X', Y') + \lambda \mathcal{L}_d(D, D', S), \quad (1)$$

Where $\mathcal{L}_p(X, Y)$ and $\mathcal{L}_p(X', Y')$ are the keypoint detector loss function for the original and a wrapped image respectively, defined as follows:

$$\mathcal{L}_p(X, Y) = \frac{1}{H_c W_c} \sum_{\substack{h=1 \\ w=1}}^{H_c W_c} l_p(x_{hw}; y_{hw}), \quad (2)$$

with:

$$l_p (x_{hw}; y) = -\log\left(\frac{\exp(x_{hwy})}{\sum_{k=1}^{65} \exp(x_{hwk})}\right), \quad (3)$$

where $x_{hw} \in X$ are pixel cells of the input image while $y_{hw} \in Y$ the corresponding labels.

The loss function of the descriptor can be defined below:

$$\mathcal{L}_d(D, D', S) = \frac{1}{(H_c W_c)^2} \sum_{\substack{h=1 \\ w=1}}^{H_c W_c} \sum_{\substack{h'=1 \\ w'=1}}^{H_c W_c} l_d(\|d_{hw}\|_2, \|d_{h'w'}\|_2; s_{hwh'w'}), \quad (4)$$

Where: $\|d_{hw}\|_2$ and $\|d_{h'w'}\|_2 \in D$ are the normalized descriptor cells from the original and wrapped image respectively while $s_{hwh'w'}$ is a binary variable which presents the homography correspondence between (h, w) and (h’, w’) cells.

Moreover, the parameter λ_d was added, aiming to reinforce the balance between negative and positive correspondences while the hinge loss is used (5):

$$l_d(d, d'; s) = \lambda_d * s * \max(0, m_p - d^T d') + \\ + (1 - s) * \max(0, d^T d' - m_n), (5)$$

where m_p and m_n are the positive and negative margins (Rosasco, 2004).

It’s worth noting that, in the original SuperPoint, the descriptor cells ($\|d_{hw}\|_2, \|d_{h'w'}\|_2$) are not normalized. However, it was observed that the normalized descriptors, combined with tuning the factor λ (eq. 1) and the weighting term λ_d (eq. 5) accordingly, produced more accurate results in unstructured environments.

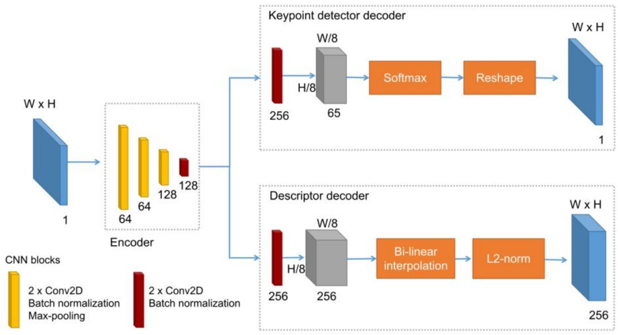


Figure 1. SuperPoint architecture

2.2 Self-supervised training of SuperPoint

The self-supervized training process of SuperPoint is conducted in several rounds aiming to increase the accuracy of feature detection. At first, the standalone keypoint detector, called MagicPoint (DeTone, 2018) is trained using a generated synthetic dataset which includes 2D geometric shapes such as lines, ellipses, triangles etc. During the training process, homographic adaptation is performed, which combines multiple random homographies of the input image and the keypoint predictions of the model, aiming to reinforce the efficiency in geometric transformations (fig 2).

After the first round of training, the trained model is used in order to extract pseudo-ground truth of the desired dataset (fig 3) while afterwards, the MagicPoint is re-trained using the desired dataset and the extracted labels while the homographic adaptation is utilized also. It's worth noting that, the MagicPoint training with the desired dataset can be repeated for two or three rounds using the optimized pseudo-ground truth each time, in order to further improve the detector's accuracy.

Finally, the SuperPoint including detector and descriptor is trained using the desired dataset and the optimized pseudo-ground truth (fig. 4).

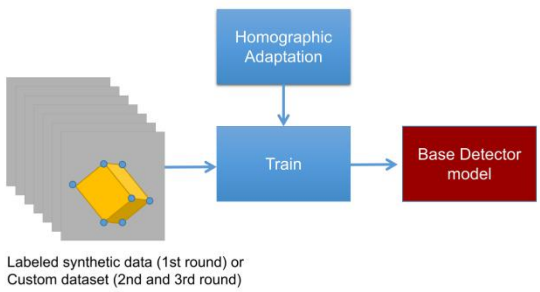


Figure 2. MagicPoint training using homographic adaptation

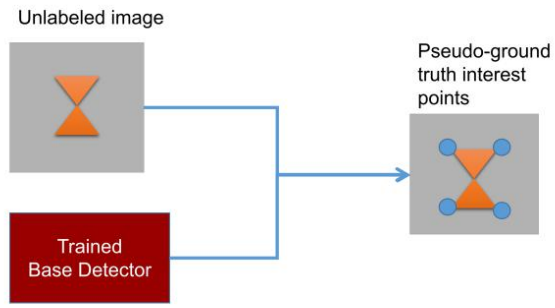


Figure 3. Pseudo-ground truth prediction based on the trained model

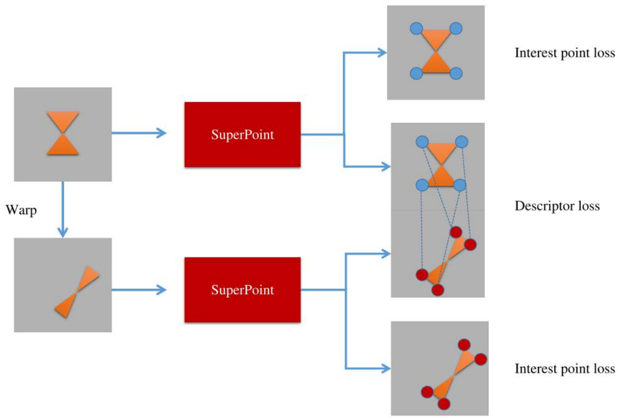


Figure 4. SuperPoint training and fine-tuning

3. IMPLEMENTATION AND RESULTS

In this section, the implementation and training procedure of the SuperPoint architecture are presented, while afterwards the evaluation and results of the extracted models are described.

3.1 SuperPoint implementation and training

SuperPoint was implemented using the TensorFlow (Abadi. 2015) deep learning platform and trained utilizing the dataset proposed in (Petrakis & Partsinevelos 2023), aiming to increase the SuperPoint's sensitivity in planetary and unstructured scenes.

As described in section 2.1, the original SuperPoint's architecture was improved applying the following two modifications:

- The bi-linear interpolation is utilized for feature maps reconstruction in full-sized images instead of bi-cubic interpolation, used by the original SuperPoint
- In the loss function of keypoint description, the descriptors of initial and wrapped images are L2 normalized while tuning the weighting parameters including λ and λ_d accordingly

During the experimentation, three SuperPoint models were produced following the training approaches presented below:

- The original SuperPoint was trained from scratch, using the aforementioned dataset aiming to focus on planetary environments

- The original SuperPoint was trained using the aforementioned datastet, based on the weights extracted by the training of SuperPoint with COCO (Lin *et al.* 2014) dataset (fine-tuning). This model aims to combine the general-purpose knowledge, with the specialized knowledge for unstructured environments acquired by the main training process
- The optimized SuperPoint trained from scratch, using the aforementioned dataset, aiming to focus on planetary environments

Both, original and optimized SuperPoint models were trained under the same parameterization. For each model, the MagicPoint which is the standalone detector of SuperPoint, was trained for three rounds applying 18 000 iterations with batch size equal to 32 and homographic adaptation enabled. Subsequently, SuperPoint was trained for 250 000 iterations with batch size equal to 2 with homographic adaptation disabled due to high demands on computing resources. The Adam optimizer with default learning rate equal to 0.001 were utilized while the image input size that was used is 240 x 320 in grayscale.

Before each round of training, the weights from the last round are used to extract the pseudo-ground truth of the dataset which is subsequently used in the next round of training. It's worth noting that in the first round, the pseudo-ground truth is extracted using the weights based on a MagicPoint model, trained with the synthetic shapes dataset.
Regarding the computing resources, an Intel i7-4771 CPU with 3.50GHz × 8 combined with an NVIDIA GeForce GTX 1080 Ti GPU were utilized while an external hard drive of 3.5 inches and a size of 4TB was used for retrieving and storing data during the training.

3.2 Evaluation and results of SuperPoint models

In this section, implemented SuperPoint models focused on unstructured environments, are evaluated in terms of keypoint detection and description, compared with well-known and widely used algorithms and the pre-trained SuperPoint model.

The evaluation is conducted using the benchmark dataset proposed in (Petrakis & Partsinevelos 2023) designed for planetary and unstructured scenes while the repeatability and homography estimation metrics are utilized for the evaluation of keypoint detection and description respectively.

Regarding the evaluation of keypoint detection, the produced models are compared with the algorithms SHI (Shi & Tomasi 1993), Harris (Harris & Stephens 1988), and FAST (Rosten & Drummond 2006) implemented with OpenCV library (Bradski 2000) and the original SuperPoint model, trained with 80 000 general-purpose images from COCO dataset. The repeatability metric, which determines the efficiency of the model to detect the same keypoints in different image representations of the same scene, was estimated using 300 detected points as the maximum limit and threshold of correctness ε=3 pixels (table 1).

Keypoint detectors	Rep. (i)	Rep. (v)
FAST	0.72	0.61
Harris	0.75	0.73
SHI	0.74	0.61
Original SuperPoint (Pre-trained)	0.85	0.63
Original SuperPoint (Trained from scratch)	0.83	0.65
Original SuperPoint (Fine-tuning)	0.83	0.65
Optimized SuperPoint (Trained from scratch)	0.82	0.66

Table 1. Evaluation of keypoint detectors based on illumination (i) and viewpoint (v) changes in planetary and unstructured environments, using repeatability metric with ε=3

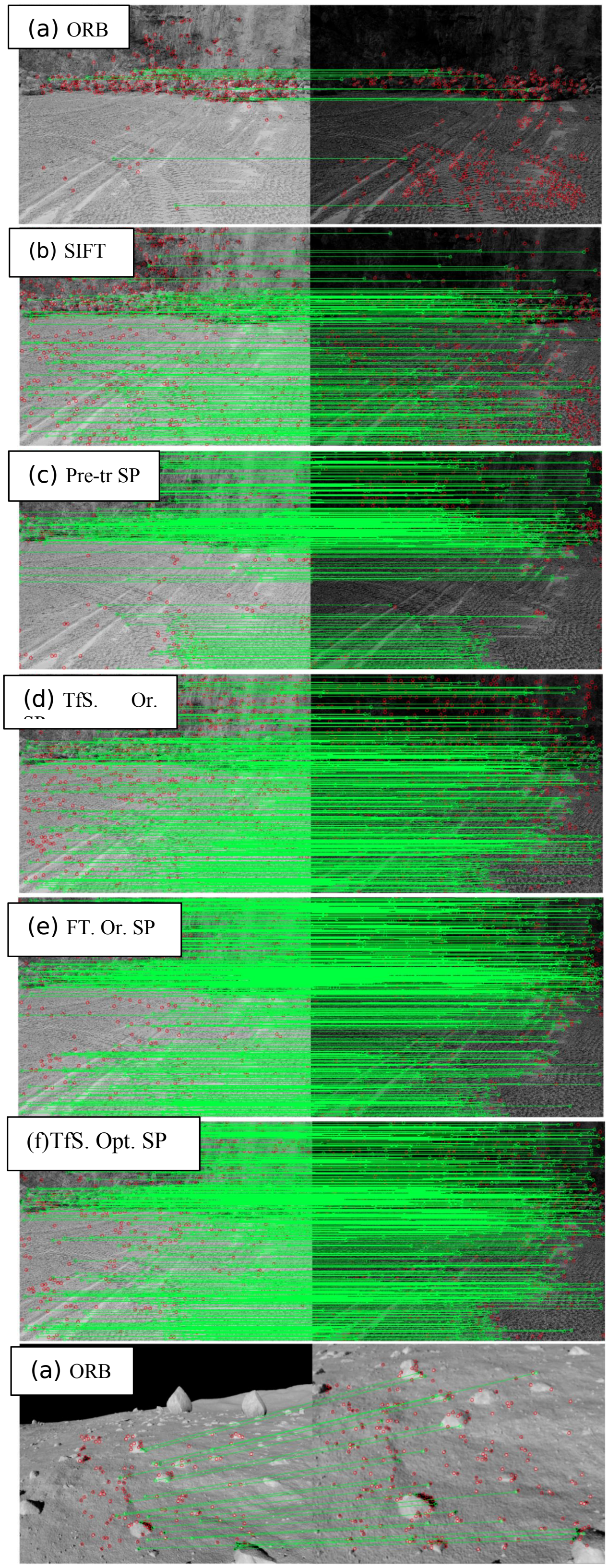
Descriptors	H.E (i)	H.E (v)
ORB	0.82	0.53
SIFT	0.97	0.96
Original SuperPoint (Pre-trained)	0.98	0.81
Original SuperPoint (Trained from scratch)	0.99	0.85
Original SuperPoint (Fine-tuning)	0.98	0.84
Optimized SuperPoint (Trained from scratch)	0.99	0.87

Table 2 Evaluation of keypoint descriptors based on illumination (i) and viewpoint (v) changes in planetary and unstructured environments, using homography estimation (H.E) with ε=3

As presented in table 1, the optimized and original SuperPoint models trained and fine-tuned with the proposed dataset, provided similar repeatability of 0.82 and 0.83 respectively in terms of illumination changes, outperforming the SHI, Harris and FAST detectors, while the pre-trained SuperPoint model achieves the highest repeatability equal to 0.85. Instead, the optimized SuperPoint model outperforms SHI, FAST and all the original SuperPoint models, (the pre-trained model and trained from scratch with the proposed dataset) in terms of viewpoint changes, achieving a repeatability score equal to 0.66. It's worth noting that Harris detector provides the highest repeatability in terms of viewpoint changes, equal to 0.73.

As presented in table 2, the optimized and original SuperPoint models provide the highest homography estimation in terms of illumination changes (0.99) outperforming the descriptors ORB, SIFT and the original pre-trained and fine-tuned SuperPoint models. In terms of viewpoint changes, the SIFT algorithm provides high accuracy in a level of 0.95 while the optimized SuperPoint model, achieves homography estimation equal to 0.87, outperforming ORB and all the original SuperPoint models (the trained and fine-tuned with the proposed dataset models and the pre-trained model).

Qualitative results in keypoint matching between images with different illumination or viewpoint are depicted in figure 5.



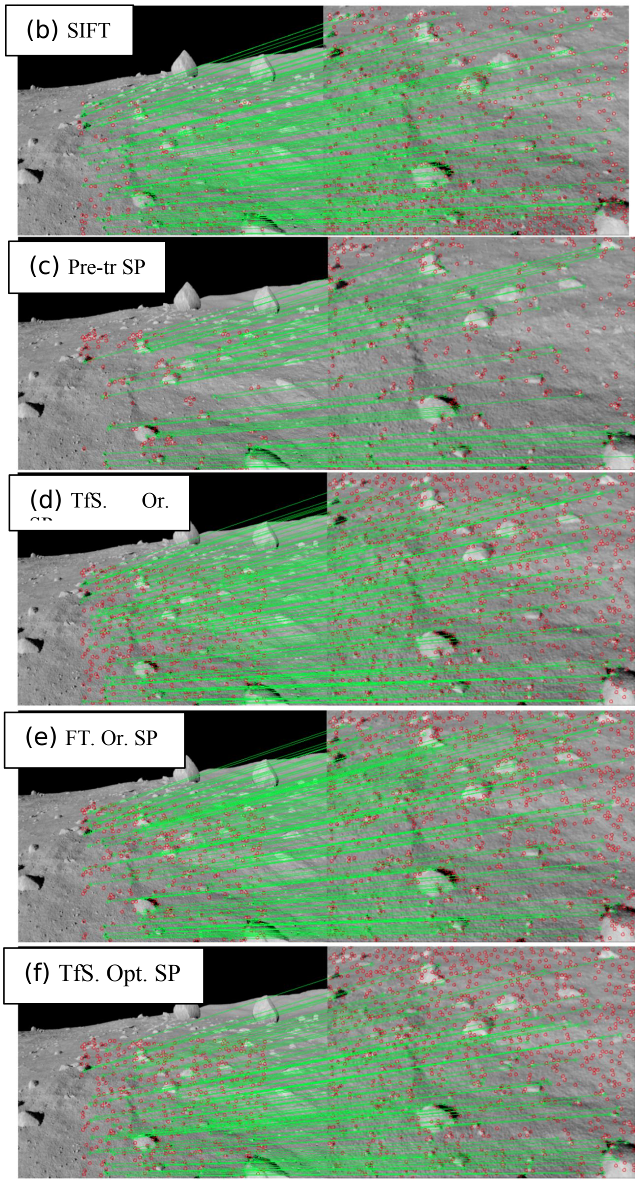


Figure 5 a - f: Keypoint matches in two images from an earthy scene in different levels of illumination: (a) ORB, (b) SIFT, (c) Pre-trained SuperPoint, (d) original SuperPoint, trained from scratch, (e) original fine-tuned SuperPoint, (f) optimized SuperPoint, trained from scratch. g-l: Keypoint matches in two images from the same lunar scene in different viewpoints: (g) ORB, (h) SIFT, (i) Pre-trained SuperPoint, (j) original SuperPoint, trained from scratch, (k) original fine-tuned SuperPoint, (l) optimized SuperPoint, trained from scratch

As observed in figure 5, the trained from scratch original and optimized SuperPoint models, using the aforementioned dataset, provide high accuracy and sensitivity in feature-poor scenes with illumination and viewpoint changes, outperforming the handcrafted algorithms and the pre-trained SuperPoint model. The fine-tuned SuperPoint, provides refined results compared with the pre-trained SuperPoint but it is not as accurate as the trained from scratch models.

4. DISCUSSION

Regarding the evaluation of keypoint detectors in terms of illumination changes, although the original and optimized models perform respectable results outperforming the handcrafted algorithms, the pre-trained SuperPoint provides the highest score in repeatability. This is reasonable, since the COCO dataset which has been utilized for the pre-trained SuperPoint model, includes thousands of images with increased variance in lighting conditions, instead of the proposed dataset which includes limited variance in illumination changes. On the contrary, the optimized SuperPoint model achieves the highest repeatability in terms of viewpoint changes, proving that enriching the dataset with high variance in illumination changes, the overall accuracy of the optimized SuperPoint will be further enhanced. Concerning the evaluation of descriptors, the optimized SuperPoint model outperforms all the original SuperPoint models and ORB algorithm in illumination and viewpoint changes while SIFT achieves the highest score in overall homography estimation.

In figure 6, the progress of the SuperPoint's learning process is presented through the visualization of detected features in a scene from Mars. Initially, in fig. 6a the features are detected using the MagicPoint (the standalone detector of SuperPoint) model trained with the synthetic shapes dataset, while afterwards the results of the MagicPoint models produced by two rounds of MagicPoint training with the proposed dataset (fig. 5b, 5c), prove the increased sensitivity in feature-poor planetary scenes. Finally, in fig 6d, the superiority of the final SuperPoint model is presented through the multiple detected features which describe the content of each scene with quite higher detail than the aforementioned MagicPoint models.

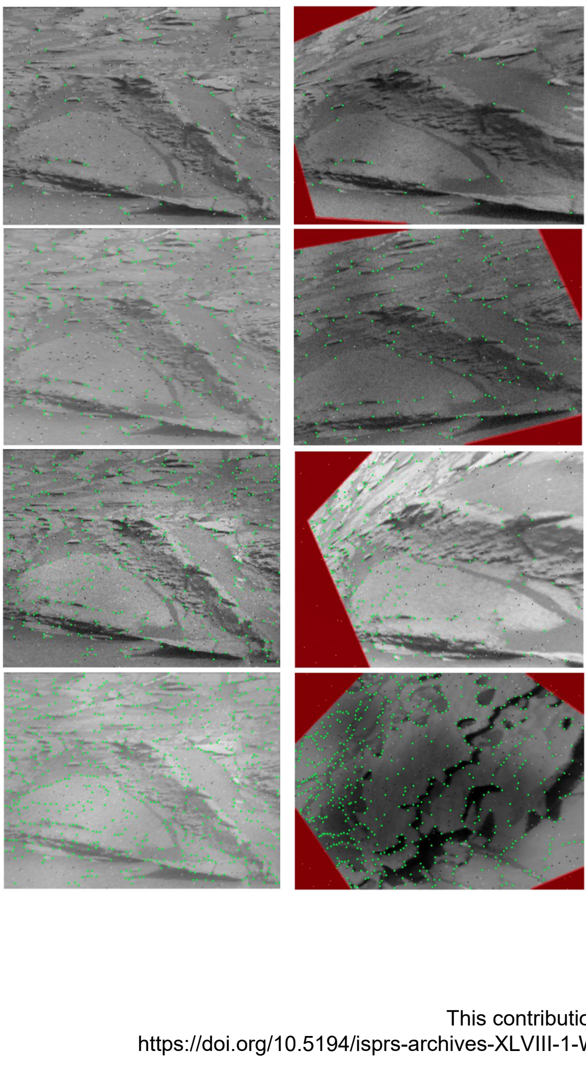


Figure 6. (a) MagicPoint model trained with synthetic shapes dataset (b) first round of MagicPoint training with the proposed dataset, (c) second round of MagicPoint training with the proposed dataset, (d SuperPoint model, trained after two rounds of MagicPoint training

It's worth mentioning that most of the studies which utilize feature extractors based on deep learning, use models that have been trained with general-purpose datasets such as COCO, regardless of the environments that are applied. The superiority of SuperPoint models, trained for unstructured environments, compared with the pre-trained SuperPoint, proves that the feature extractors based on deep learning, trained for a specialized and completely different environment, are able to provide increased efficiency compared with a model trained with general-purpose datasets.

5. CONCLUSIONS

In this study, an investigation of SuperPoint architecture's efficiency in keypoint detection and description applied to unstructured and planetary scenes was conducted. Two modifications in the original SuperPoint architecture including the use of bi-linear instead of bicubic interpolation in the descriptor decoder and the normalization of the descriptors in the descriptor's loss function, were implemented, aiming to increase the accuracy of the model in unstructured environments. The original and an optimized architecture of SuperPoint, were trained with the proposed dataset, producing three different models: (a) an original SuperPoint model trained from scratch, (b) an original fine-tuned SuperPoint model, (c) an optimized SuperPoint model, trained from scratch with the same parametarization as the corresponding original model. The models were evaluated using the designed benchmark dataset while the repeatability and homography estimation metrics were utilized in order to evaluate the produced models and compared with the pre-trained SuperPoint model, trained with COCO dataset and several popular keypoint detectors and descriptors.
The results determined a scaleable potential of deep learning in unstructured environments while the optimized SuperPoint model, provided satisfactory accuracy compared with the pre-trained and fine-tuned SuperPoint.

However, the lack of the utilized dataset with high variance in illumination changes is one of the main limitations of the optimized and original SuperPoint models, since it is the main reason of providing slightly lower repeatability in keypoint detection compared with the pre-trained SuperPoint. Moreover, although the optimized model outperforms all the original SuperPoint models, SIFT algorithm provides the highest accuracy in viewpoint changes.

Thus, the future work of the optimized model can be focused on two main improvements. At first, the proposed dataset could be enriched including real or artificial images with intense lighting changes, while the difficulty of homographic adaptation during the training process could be increased, through more examples with highly transformed representations. The aforementioned improvements, will further escalate the efficiency of the model in illumination and viewpoint changes, providing more robust results in both illumination and viewpoint changes.

As a conclusion, the optimized SuperPoint model, is a promising solution for accurate keypoint detection and description in unstructured and planetary scenes, which could be an inspiration for the computer vision community, increasing the potential for accurate autonomous navigation in completely unknown and unstructured scenes.

ACKNOWLEDGEMENTS

The implementation of the doctoral thesis was co-financed by Greece and the European Union (European Social Fund-ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning» in the context of the Act “Enhancing Human Resources Research Potential by undertaking a Doctoral Research” Sub-action 2: IKY Scholarship Programme for PhD candidates in the Greek Universities.

REFERENCES

Abadi M, Agarwal A, Barham P et al, TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org

Balntas, V., Lenc, K., Vedaldi, A., Mikolajczyk, K.: 2017, HPatches: A Benchmark and Evaluation of Handcrafted and Learned Local Descriptors, Proceedings of CVPR, pp. 5173-5182

Bojanić, D., Bartol, K., Pribanić, T., Petković, T., Donoso, Y.D, Mas, J. S: 2019, On the Comparison of Classic and Deep Keypoint Detector and Descriptor Methods, 11th ISPA, Dubrovnik, Croatia pp. 64-69

Bradski G, The OpenCV Library. Dr Dobb's Journal of Software Tools. 2000

Christiansen, P., Kragh, M., Brodskiy, Y., Karstoft, H., 2019: UnsuperPoint: End-to-end Unsupervised Interest Point Detector and Descriptor, arXiv

DeTone, D., Malisiewicz, T., Rabinovich, A., 2018: SuperPoint: Self-Supervised Interest Point Detection and Description, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 224-236

Dusmanu, M., Rocco, I., Pajdla, T., Pollefeys, M., Sivic, J., Torii, A., Sattler, T.: 2019, D2-Net: A Trainable CNN for Joint Detection and Description of Local Features, CVPR

Furgale, T., Carle, P., Enright, J., Barfoot, D., The Devon Island Rover Navigation Dataset, International Journal of Robotics Research, 2012

Giubilato, R., Stürzl, W., Wedler, A., Triebel, R.: 2022, Challenges of SLAM in Extremely Unstructured Environments: The DLR Planetary Stereo,Solid-State LiDAR, Inertial Dataset, IEEE Robotics and automation letters vol.7

Guastella, DC., Muscato, G.: 2021, Learning-Based Methods of Perception and Navigation for Ground Vehicles in Unstructured Environments: A Review. MDPI, Sensors

Harris, C., Stephens, M., A. 1988: Combined Corner and Edge Detector, Alvey Vision Conference. Vol. 15

Hewitt, R., Boukas, E., Azkarate, M., Pagnamenta, M., Marshall, J., Gasteratos, A., Visentin, G.:2018 The Katwijk beach planetary roverdataset, The international Journal of Robotics research

Isık, S., Ozkan, K: 2015, A Comparative Evaluation of Well-known Feature Detectors and Descriptors . International Journal of Applied Mathematics Electronics and Computers

Lin, T., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, L.: 2014, Microsoft COCO: Common Objects in Context. ECCV, Springer

Liu, Y., Li, J., Huang, K., Xiangting, L., Xiuyuan, Q., Chang, L., Long, Y., Zhou, J.: 2022, MobileSP: An FPGA-Based Real-Time Keypoint Extraction Hardware Accelerator for Mobile VSLAM, in IEEE Transactions on Circuits and Systems I: Regular Papers2

Lowe, D. 2004: Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision. 60 (2): 91–110

Ma, J., Jiang, X., Fan, A., Jiang, J., Yan, J., 2021: Image Matching from Handcrafted to Deep Features: A Survey. Int J Comput Vis 129, 23–79

Meyer, L., Smísek, M., Villacampa, F, Maza, L., Medina, D., Schuster, M, Steidle, F., Vayugundla, M., Muller M., Rebele, B., Wedler, A., Triebel, R.:2022, The MADMAX data set for visual-inertial rover navigation on Mars, J Field Robotics

Nixon, M., Aguado, A., 2019: Feature Extraction and Image Processing for Computer Vision; Academic Press: Cambridge, MA, USA

Noh, H., Araujo, A., Sim, J., Weyand, T., Han, B.: 2017, Large-scale image retrieval with attentive deep local features. In Proceedings of CVPR

Ono, Y., Trulls, E., Fua, P., Yi, K., 2018: LF-Net: Learning Local Features from Images. On NIPS

Petrakis, G., Partsinevelos, P.: 2023, Keypoint detection and description in unstructured environments through deep learning, Drones, MDPI (submitted June 23, under review)

Ronneberger, O., Fischer, P., Brox, T.: 2015: U-Net: Convolutional Networks for Biomedical Image Segmentation, MICCAI 2015

Rosasco, L., Vito, E., Caponnetto, A., Piana, M., Verri, A.: 2004, Are loss functions all the same? Neural Computation. 16

Rosten, E., Drummond, T. 2006: Machine Learning for High-speed Corner Detection, Computer Vision – ECCV Lecture Notes in Computer Science. Vol. 3951. pp. 430–443

Rublee, E., Rabaud, V., Konolige, K., Bradski, G. 2011: ORB: An efficient alternative to SIFT or SURF, ICCV, Barcelona, Spain pp. 2564-2571

Sattler, T., Weyand, T., Leibe, B., Kobbelt, L.:2008, Image retrieval for image-based localization revisited. In proceedings of CVPR

Shi J., Tomasi, C. 1993: Good features to track. Technical report, Cornell University

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A., Bishop, R., Rueckert, D., Wang, Z.: 2016, Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network, CVPR

Simonyan. K., Zisserman, A.: 2014, A Very deep convolutional networks for large-scale image recognition, ArXiv 201