

University of Groningen

Data-efficient representation learning for visual place recognition

Leyva Vallina, María

DOI:
[10.33612/diss.736449452](https://doi.org/10.33612/diss.736449452)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2023

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Leyva Vallina, M. (2023). *Data-efficient representation learning for visual place recognition*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.736449452>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Bibliography

- Angelina Uy, M. and Hee Lee, G.: 2018, Pointnetvlad: Deep point cloud based retrieval for large-scale place recognition, *CVPR*, pp. 4470–4479.
- Arandjelovic, R., Gronat, P., Torii, A., Pajdla, T. and Sivic, J.: 2016, Netvlad: Cnn architecture for weakly supervised place recognition, *CVPR*, pp. 5297–5307.
- Arandjelovic, R. and Zisserman, A.: 2013, All about vlad, *CVPR*, pp. 1578–1585.
- Babenko, A. and Lempitsky, V.: 2015, Aggregating local deep features for image retrieval, *CVPR*, pp. 1269–1277.
- Bac, C. W., van Henten, E. J., Hemming, J. and Edan, Y.: 2014, Harvesting robots for high-value crops: State-of-the-art review and challenges ahead, *Journal of Field Robotics* **31**(6), 888–911.
- Badino, H., Huber, D. and Kanade, T.: 2011a, The CMU Visual Localization Data Set, <http://3dvis.ri.cmu.edu/data-sets/localization>.
- Badino, H., Huber, D. and Kanade, T.: 2011b, Visual topometric localization, *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pp. 794–799.
- Badrinarayanan, V., Kendall, A. and Cipolla, R.: 2017, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**(12), 2481–2495.
- Balntas, V., Li, S. and Prisacariu, V.: 2018, Relocnet: Continuous metric learning relocalisation using neural nets, *ECCV*, pp. 751–767.
- Baslamisli, A. S., Groenestege, T. T., Das, P., Le, H.-A., Karaoglu, S. and Gevers, T.: 2018, Joint learning of intrinsic images and semantic segmentation, *ECCV*, pp. 286–302.
- Berton, G., Masone, C. and Caputo, B.: 2022, Rethinking visual geo-localization for large-scale applications, *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4868–4878.

- Bromley, J., Guyon, I., LeCun, Y., Säckinger, E. and Shah, R.: 1994, Signature verification using a "siamese" time delay neural network, *NeurIPS*, pp. 737–744.
- Cao, B., Araujo, A. and Sim, J.: 2020, Unifying deep local and global features for image search, in A. Vedaldi, H. Bischof, T. Brox and J.-M. Frahm (eds), *ECCV 2020*, pp. 726–743.
- Caruana, R.: 1997, Multitask learning, *Machine learning* **28**(1), 41–75.
- Chang, Y.-T., Wang, Q., Hung, W.-C., Piramuthu, R., Tsai, Y.-H. and Yang, M.-H.: 2020, Weakly-supervised semantic segmentation via sub-category exploration, *CVPR*, pp. 8991–9000.
- Chen, T., Kornblith, S., Norouzi, M. and Hinton, G.: 2020, A simple framework for contrastive learning of visual representations, *International conference on machine learning*, PMLR, pp. 1597–1607.
- Chen, X. and He, K.: 2021, Exploring simple siamese representation learning, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 15750–15758.
- Chen, Z., Lam, O., Jacobson, A. and Milford, M.: 2014, Convolutional neural network-based place recognition, *ACRA* .
- Chopra, S., Hadsell, R. and LeCun, Y.: 2005, Learning a similarity metric discriminatively, with application to face verification, *CVPR*, Vol. 1, IEEE, pp. 539–546.
- Cummins, M. and Newman, P.: 2008, Fab-map: Probabilistic localization and mapping in the space of appearance, *The International Journal of Robotics Research* **27**(6), 647–665.
- Cummins, M. and Newman, P.: 2009, Highly scalable appearance-only slam-fab-map 2.0., *Robotics: Science and Systems*, Vol. 5, p. 17.
- Dalal, N. and Triggs, B.: 2005, Histograms of oriented gradients for human detection, *IEEE CVPR*, Vol. 1, pp. 886–893.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L.: 2009, Imagenet: A large-scale hierarchical image database, *CVPR*, pp. 248–255.
- Ding, M., Wang, Z., Sun, J., Shi, J. and Luo, P.: 2019, Camnet: Coarse-to-fine retrieval for camera re-localization, *ICCV*, pp. 2871–2880.
- Doan, D., Latif, Y., Chin, T., Liu, Y., Do, T. and Reid, I.: 2019, Scalable place recognition under appearance change for autonomous driving, *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9318–9327.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. and Houshy, N.: 2021, An image is worth 16x16 words: Transformers for image recognition at scale, *International Conference on Learning Representations*.
URL: <https://openreview.net/forum?id=YicbFdNTTy>

- Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J. and Zisserman, A.: 2015, The pascal visual object classes challenge: A retrospective, *Int. J. Comput. Vis.* **111**(1), 98–136.
- Furgale, P., Rehder, J. and Siegwart, R.: 2013, Unified temporal and spatial calibration for multi-sensor systems, *IEEE IROS*, pp. 1280–1286.
- Galvez-López, D. and Tardos, J. D.: 2012, Bags of binary words for fast place recognition in image sequences, *IEEE Transactions on Robotics* **28**(5), 1188–1197.
- Geiger, A., Lenz, P., Stiller, C. and Urtasun, R.: 2013, Vision meets robotics: The kitti dataset, *Int. J. Robot. Res.* **32**(11), 1231–1237.
- Gordo, A., Almazán, J., Revaud, J. and Larlus, D.: 2017, End-to-end learning of deep visual representations for image retrieval, *International Journal of Computer Vision* **124**(2), 237–254.
- Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z., Gheshlaghi Azar, M. et al.: 2020, Bootstrap your own latent—a new approach to self-supervised learning, *Advances in neural information processing systems* **33**, 21271–21284.
- Hadsell, R., Chopra, S. and LeCun, Y.: 2006, Dimensionality reduction by learning an invariant mapping, *CVPR*, Vol. 2, IEEE, pp. 1735–1742.
- Hausler, S., Garg, S., Xu, M., Milford, M. and Fischer, T.: 2021, Patch-netvlad: Multi-scale fusion of locally-global descriptors for place recognition, *CVPR*, pp. 14141–14152.
- He, K., Zhang, X., Ren, S. and Sun, J.: 2016, Deep residual learning for image recognition, *CVPR*, pp. 770–778.
- Honegger, D., Sattler, T. and Pollefeys, M.: 2017, Embedded real-time multi-baseline stereo, *IEEE ICRA*.
- Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K. Q.: 2017, Densely connected convolutional networks, *CVPR*, pp. 4700–4708.
- Jégou, H., Douze, M., Schmid, C. and Pérez, P.: 2010, Aggregating local descriptors into a compact image representation, *CVPR*, IEEE, pp. 3304–3311.
- Jegou, H., Perronnin, F., Douze, M., Sánchez, J., Perez, P. and Schmid, C.: 2011, Aggregating local image descriptors into compact codes, *TPAMI* **34**(9), 1704–1716.
- Kendall, A. and Cipolla, R.: 2016, Modelling uncertainty in deep learning for camera relocalization, *IEEE ICRA*.
- Kendall, A. and Cipolla, R.: 2017, Geometric loss functions for camera pose regression with deep learning, *CVPR*, pp. 5974–5983.
- Kendall, A., Gal, Y. and Cipolla, R.: 2018, Multi-task learning using uncertainty to weigh losses for scene geometry and semantics, *CVPR*, pp. 7482–7491.

- Kendall, A., Grimes, M. and Cipolla, R.: 2015a, Posenet: A convolutional network for real-time 6-dof camera relocalization, *IEEE ICCV*, pp. 2938–2946.
- Kendall, A., Grimes, M. and Cipolla, R.: 2015b, Posenet: A convolutional network for real-time 6-dof camera relocalization, *CVPR*, pp. 2938–2946.
- Kim, S., Kim, D., Cho, M. and Kwak, S.: 2021, Embedding transfer with label relaxation for improved metric learning, *CVPR*, pp. 3967–3976.
- Kim, S., Seo, M., Laptev, I., Cho, M. and Kwak, S.: 2019, Deep metric learning beyond binary supervision, *CVPR*, pp. 2288–2297.
- Košecká, J., Li, F. and Yang, X.: 2005, Global localization and relative positioning based on scale-invariant keypoints, *Robotics and Autonomous Systems* **52**(1), 27–38.
- Krizhevsky, A., Sutskever, I. and Hinton, G. E.: 2012, Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, pp. 1097–1105.
- Le, H., Baslamisli, A. S., Mensink, T. and Gevers, T.: 2018, Three for one and one for three: Flow, segmentation, and surface normals, *BMVC*, p. 145.
- Leyva-Vallina, M., Strisciuglio, N., Lopez-Antequera, M., Tylecek, R., Blaiçh, M. and Petkov, N.: 2019, Tb-places: A data set for visual place recognition in garden environments, *IEEE Access*.
- Leyva-Vallina, M., Strisciuglio, N. and Petkov, N.: 2023, Generalized contrastive optimization of siamese networks for place recognition, *International Conference on Learning Representations*.
URL: <https://arxiv.org/abs/2303.11739>
- LeyvaVallina, M., Strisciuglio, N. and Petkov, N.: 2019, Place recognition in gardens by learning visual representations: data set and benchmark analysis, *CAIP*, Springer, pp. 324–335.
- Li, Y., Snavely, N. and Huttenlocher, D. P.: 2010, Location recognition using prioritized feature matching, *ECCV*, Springer, pp. 791–804.
- Li, Y., Snavely, N., Huttenlocher, D. P. and Fua, P.: 2016, Worldwide pose estimation using 3d point clouds, *Large-Scale Visual Geo-Localization*, Springer, pp. 147–163.
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C. L.: 2014, Microsoft coco: Common objects in context, *ECCV*, Springer, pp. 740–755.
- Liu, L., Li, H. and Dai, Y.: 2019, Stochastic attraction-repulsion embedding for large scale image localization, *CVPR*, pp. 2570–2579.
- Lopez-Antequera, M., Gomez-Ojeda, R., Petkov, N. and Gonzalez-Jimenez, J.: 2017, Appearance-invariant place recognition by discriminatively training a convolutional neural network, *Pattern Recognit. Lett.* **92**, 89–95.

- Lopez-Antequera, M., Leyva-Vallina, M., Strisciuglio, N. and Petkov, N.: 2019, Place and object recognition by cnn-based cosfire filters, *IEEE Access* **7**, 66157–66166.
- Lowry, S., Sünderhauf, N., Newman, P., Leonard, J. J., Cox, D., Corke, P. and Milford, M. J.: 2016, Visual place recognition: A survey, *IEEE Transactions on Robotics* **32**(1), 1–19.
- Maddern, W., Pascoe, G., Linegar, C. and Newman, P.: 2017, 1 year, 1000 km: The oxford robotcar dataset, *Int. J. Robot. Res.* **36**(1), 3–15.
- Masone, C. and Caputo, B.: 2021, A survey on deep visual place recognition, *IEEE Access* pp. 1–1.
- McManus, C., Churchill, W., Maddern, W., Stewart, A. D. and Newman, P.: 2014, Shady dealings: Robust, long-term visual localisation using illumination invariance, *IEEE ICRA*, pp. 901–906.
- Milford, M. J. and Wyeth, G. F.: 2012, Seqslam: Visual route-based navigation for sunny summer days and stormy winter nights, *ICRA*, pp. 1643–1649.
- Naha, S., Xiao, Q., Banik, P., Alimoor Reza, M. and Crandall, D. J.: 2020, Pose-guided knowledge transfer for object part segmentation, *CVPR Workshops*, pp. 906–907.
- Newman, P. and Ho, K.: 2005, Slam-loop closing with visually salient features, *ICRA*, pp. 635–642.
- Ohi, N., Lassak, K., Watson, R., Strader, J., Du, Y., Yang, C., Hedrick, G., Nguyen, J., Harper, S., Reynolds, D. et al.: 2018, Design of an autonomous precision pollination robot, *IEEE IROS*.
- Perronnin, F., Liu, Y., Sánchez, J. and Poirier, H.: 2010, Large-scale image retrieval with compressed fisher vectors, *CVPR, IEEE*, pp. 3384–3391.
- Philbin, J., Chum, O., Isard, M., Sivic, J. and Zisserman, A.: 2007, Object retrieval with large vocabularies and fast spatial matching, *CVPR, IEEE*, pp. 1–8.
- Philbin, J., Chum, O., Isard, M., Sivic, J. and Zisserman, A.: 2008, Lost in quantization: Improving particular object retrieval in large scale image databases, *CVPR*, pp. 1–8.
- Pion, N., Humenberger, M., Csurka, G., Cabon, Y. and Sattler, T.: 2020a, Benchmarking image retrieval for visual localization, *International Conference on 3D Vision*.
- Pion, N., Humenberger, M., Csurka, G., Cabon, Y. and Sattler, T.: 2020b, Benchmarking image retrieval for visual localization, *International Conference on 3D Vision*.
- Qi, C. R., Su, H., Mo, K. and Guibas, L. J.: 2017, Pointnet: Deep learning on point sets for 3d classification and segmentation, *CVPR*, pp. 652–660.
- Radenović, F., Tolias, G. and Chum, O.: 2016, Cnn image retrieval learns from bow: Unsupervised fine-tuning with hard examples, *ECCV, Springer*, pp. 3–20.

- Radenović, F., Tolias, G. and Chum, O.: 2018, Fine-tuning cnn image retrieval with no human annotation, *TPAMI* **41**(7), 1655–1668.
- Revaud, J., Almazán, J., Rezende, R. S. and Souza, C. R. d.: 2019, Learning with average precision: Training image retrieval with a listwise loss, *ICCV*, pp. 5107–5116.
- Sarlin, P.-E., DeTone, D., Malisiewicz, T. and Rabinovich, A.: 2020, SuperGlue: Learning feature matching with graph neural networks, *CVPR*.
URL: <https://arxiv.org/abs/1911.11763>
- Sattler, T., Maddern, W., Toft, C., Torii, A., Hammarstrand, L., Stenborg, E., Safari, D., Okutomi, M., Pollefeys, M., Sivic, J. et al.: 2018, Benchmarking 6dof outdoor visual localization in changing conditions, *CVPR*, pp. 8601–8610.
- Sattler, T., Weyand, T., Leibe, B. and Kobbelt, L.: 2012, Image retrieval for image-based localization revisited., *BMVC*, Vol. 1, p. 4.
- Schönberger, J. L., Pollefeys, M., Geiger, A. and Sattler, T.: 2018, Semantic visual localization, *ISPRS Journal of Photogrammetry and Remote Sensing (JPRS)* .
- Sharif Razavian, A., Azizpour, H., Sullivan, J. and Carlsson, S.: 2014, Cnn features off-the-shelf: an astounding baseline for recognition, *CVPR Workshops*, pp. 806–813.
- Shi, H., Yang, Y., Zhu, X., Liao, S., Lei, Z., Zheng, W. and Li, S. Z.: 2016, Embedding deep metric for person re-identification: A study against large variations, *ECCV*, Springer, pp. 732–748.
- Shotton, J., Glocker, B., Zach, C., Izadi, S., Criminisi, A. and Fitzgibbon, A.: 2013, Scene coordinate regression forests for camera relocalization in rgb-d images, *CVPR*, pp. 2930–2937.
- Simonyan, K. and Zisserman, A.: 2014, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556* .
- Simonyan, K. and Zisserman, A.: 2015, Very deep convolutional networks for large-scale image recognition, *ICLR* .
- Strisciuglio, N., Lopez-Antequera, M. and Petkov, N.: 2019, A push-pull layer improves robustness of convolutional neural networks, *arXiv preprint arXiv:1901.10208* .
- Strisciuglio, N., Tylecek, R., Blaich, M., Petkov, N., Bieber, P., Hemming, J., van Henten, E., Sattler, T., Pollefeys, M., Gevers, T., Brox, T. and Fisher, R. B.: 2018, Trimbot2020: an outdoor robot for automatic gardening, *ISR* pp. 1–6.
- Sun, C., Shrivastava, A., Singh, S. and Gupta, A.: 2017, Revisiting unreasonable effectiveness of data in deep learning era, *CVPR*, pp. 843–852.
- Sünderhauf, N., Neubert, P. and Protzel, P.: 2013, Are we there yet? challenging seqslam on a 3000 km journey across all four seasons, *ICRA*, p. 2013.

- Sünderhauf, N. and Protzel, P.: 2011, Brief-gist-closing the loop by simple means, *IEEE IROS*, IEEE, pp. 1234–1241.
- Sünderhauf, N., Shirazi, S., Dayoub, F., Upcroft, B. and Milford, M.: 2015a, On the performance of convnet features for place recognition, *IROS*, IEEE, pp. 4297–4304.
- Sünderhauf, N., Shirazi, S., Dayoub, F., Upcroft, B. and Milford, M.: 2015b, On the performance of convnet features for place recognition, *IROS*, pp. 4297–4304.
- Taira, H., Okutomi, M., Sattler, T., Cimpoi, M., Pollefeys, M., Sivic, J., Pajdla, T. and Torii, A.: 2019, InLoc: Indoor Visual Localization with Dense Matching and View Synthesis, *TPAMI*.
- Talavera Martinez, E., Leyva-Vallina, M., Sarker, M. M. K., Puig, D., Petkov, N. and Radeva, P.: 2019, Hierarchical approach to classify food scenes in egocentric photo-streams, *IEEE J. Biomed. Health* **24**(3), 866–877.
- Thoma, J., Paudel, D. P. and Van Gool, L.: 2020, Soft contrastive learning for visual localization, *NeurIPS*.
- Torii, A., Arandjelović, R., Sivic, J., Okutomi, M. and Pajdla, T.: 2013, 24/7 place recognition by view synthesis, *CVPR*.
- Torii, A., Sivic, J., Okutomi, M. and Pajdla, T.: 2015a, Visual place recognition with repetitive structures, *IEEE Transactions on Pattern Analysis and Machine Intelligence* pp. 2346–2359.
- Torii, A., Sivic, J., Okutomi, M. and Pajdla, T.: 2015b, Visual place recognition with repetitive structures, *TPAMI*.
- Uy, M. A. and Lee, G. H.: 2018, Pointnetvlad: Deep point cloud based retrieval for large-scale place recognition, *IEEE CVPR*.
- Van der Maaten, L. and Hinton, G.: 2008, Visualizing data using t-sne., *J. Mach. Learn. Res.* **9**(11).
- Walch, F., Hazirbas, C., Leal-Taixé, L., Sattler, T., Hilsenbeck, S. and Cremers, D.: 2017, Image-based localization using lstms for structured feature correlation, *IEEE ICCV*.
- Wald, J., Dhano, H., Navab, N. and Tombari, F.: 2020, Learning 3d semantic scene graphs from 3d indoor reconstructions, *CVPR*, pp. 3961–3970.
- Walter, A., Khanna, R., Lottes, P., Stachniss, C., Siegwart, R., Nieto, J. and Liebisch, F.: 2018, Flourish-a robotic approach for automation in crop management, *ICPA*.
- Wang, P., Yang, R., Cao, B., Xu, W. and Lin, Y.: 2018, Dels-3d: Deep localization and segmentation with a 3d semantic map, *IEEE CVPR*, pp. 5860–5869.
- Wang, R., Shen, Y., Zuo, W., Zhou, S. and Zhen, N.: 2022, Transvpr: Transformer-based place recognition with multi-level attention aggregation, *CVPR*.

- Warburg, F., Hauberg, S., López-Antequera, M., Gargallo, P., Kuang, Y. and Civera, J.: 2020, Mapillary street-level sequences: A dataset for lifelong place recognition, *CVPR*.
- Weyand, T., Araujo, A., Cao, B. and Sim, J.: 2020, Google landmarks dataset v2-a large-scale benchmark for instance-level recognition and retrieval, *CVPR*, pp. 2575–2584.
- Xie, S., Girshick, R., Dollár, P., Tu, Z. and He, K.: 2017, Aggregated residual transformations for deep neural networks, *CVPR*, pp. 1492–1500.
- Yu, F., Chen, H., Wang, X., Xian, W., Chen, Y., Liu, F., Madhavan, V. and Darrell, T.: 2020, Bdd100k: A diverse driving dataset for heterogeneous multitask learning, *CVPR*, pp. 2636–2645.
- Yu, J., Zhu, C., Zhang, J., Huang, Q. and Tao, D.: 2019, Spatial pyramid-enhanced netvlad with weighted triplet loss for place recognition, *IEEE Trans. Neural Netw. Learn. Syst.* **31**(2), 661–674.
- Zaffar, M., Garg, S., Milford, M., Kooij, J., Flynn, D., McDonald-Maier, K. and Ehsan, S.: 2021, Vpr-bench: An open-source visual place recognition evaluation framework with quantifiable viewpoint and appearance change, *International Journal of Computer Vision* **129**(7), 2136–2174.
- Zhang, P., Lan, C., Zeng, W., Xing, J., Xue, J. and Zheng, N.: 2020, Semantics-guided neural networks for efficient skeleton-based human action recognition, *CVPR*, pp. 1112–1121.
- Zhang, X., Wang, L. and Su, Y.: 2020, Visual place recognition: A survey from deep learning perspective, *Pattern Recognition* p. 107760.
- Zheng, S., Song, Y., Leung, T. and Goodfellow, I.: 2016, Improving the robustness of deep neural networks via stability training, *IEEE CVPR*, pp. 4480–4488.
- Zhou, B., Lapedriza, A., Khosla, A., Oliva, A. and Torralba, A.: 2017a, Places: A 10 million image database for scene recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Zhou, B., Lapedriza, A., Khosla, A., Oliva, A. and Torralba, A.: 2017b, Places: A 10 million image database for scene recognition, *TPAMI* **40**(6), 1452–1464.
- Zou, J. and Schiebinger, L.: 2018, Ai can be sexist and racist — it’s time to make it fair, *Nature* **559**, 324–326.