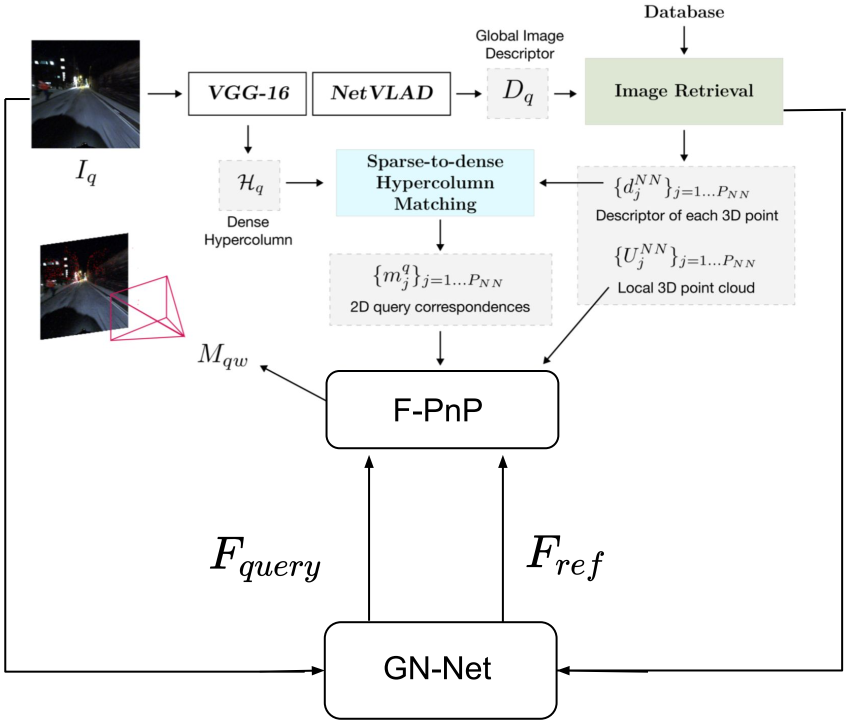
Deep Direct Sparse-to-Dense Localization

Group Member: Lixin Xue, Le Chen, Zimeng Jiang
Supervisor: Paul-Edouard Sarlin

Sparse-to-Dense Localization



$$min||F_{query}(u_q) - F_{ref}(u_r)||$$

[1]Germain et al. Sparse-To-Dense Hypercolumn Matching for Long-Term Visual Localization.

Roadmap

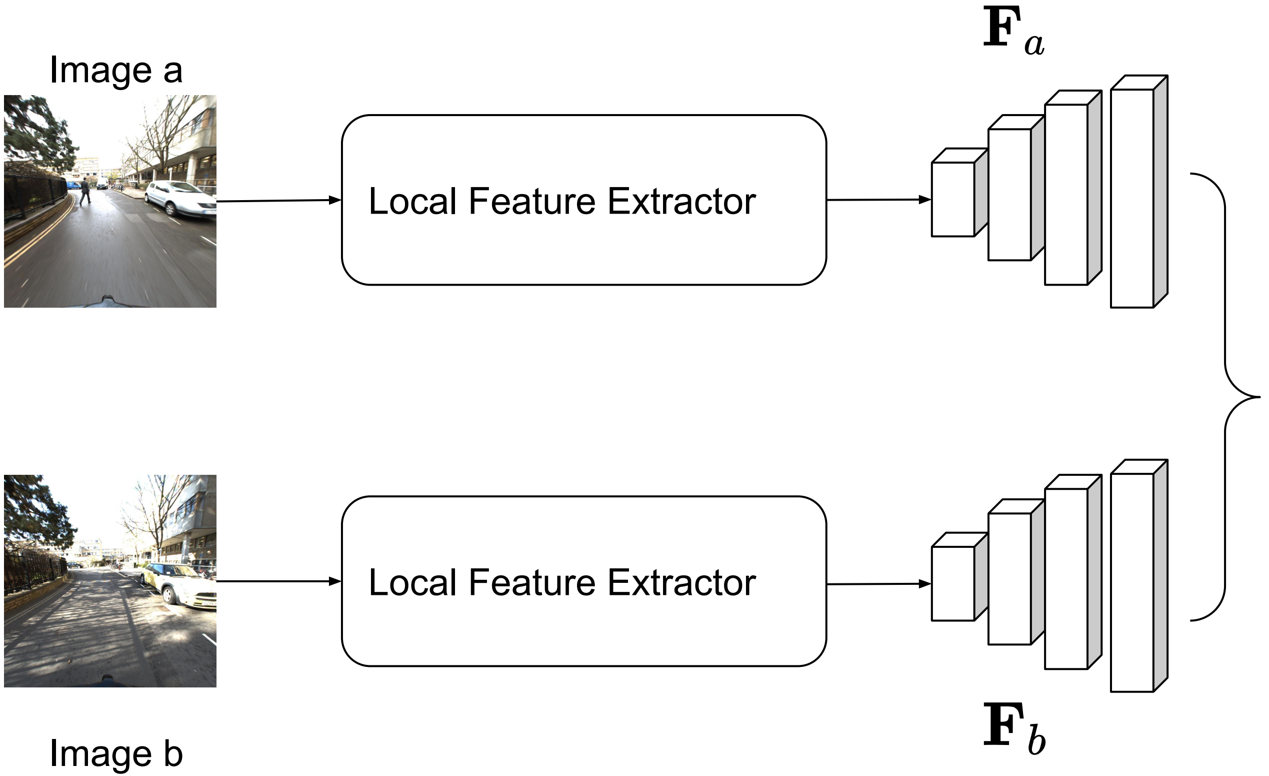
Proposal to Midterm

- Feature-metric PnP (F-PnP) implementation
- Validation on toy example

Midterm to Final

- Incorporate F-PnP into Sparse-to-dense framework
- Validation of F-PnP with retrieval features
- Gauss-Newton Net reimplementation

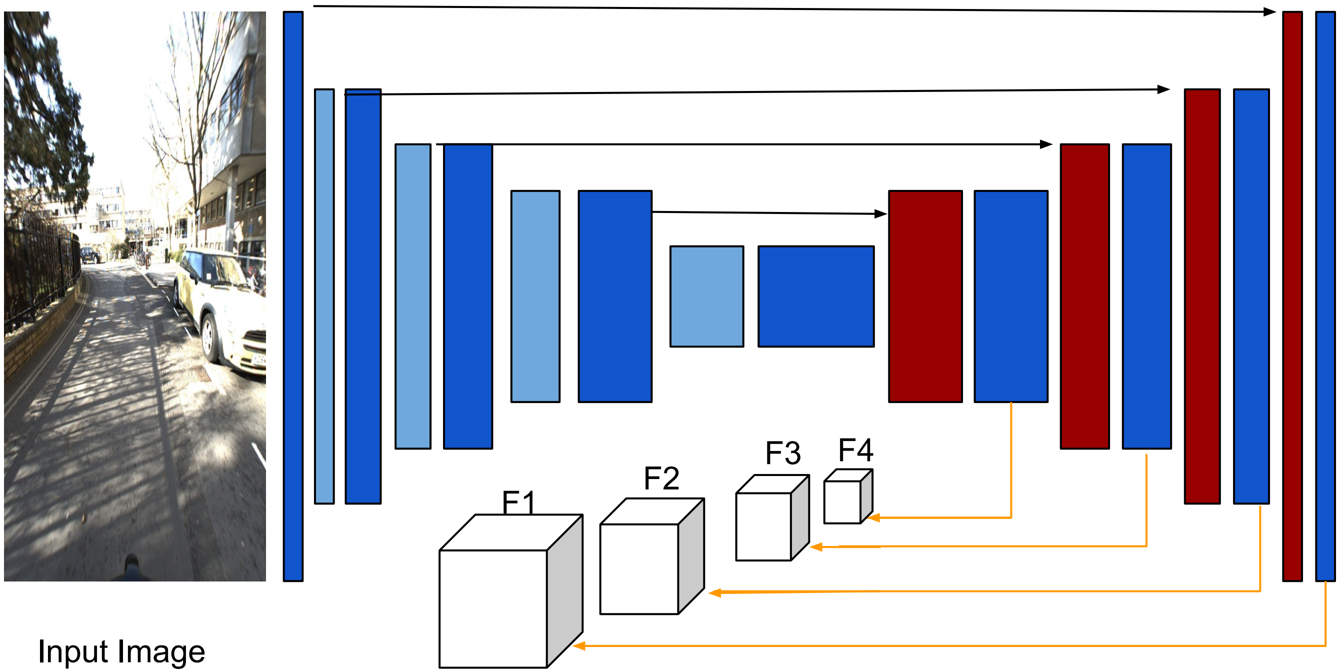
GN-Net: Learn better feature maps for feature-metric PnP



$$\mathcal{L}_{\text{total}}\left(\mathbf{F}_a, \mathbf{F}_b\right) = \lambda_{con} \sum_l \mathcal{L}_{\text{contrastive}}\left(\mathbf{F}_a, \mathbf{F}_b, l\right) + \lambda_{gn} \sum_l \mathcal{L}_{\text{gauss-newton}}\left(\mathbf{F}_a, \mathbf{F}_b, l\right)$$

[2] von Stumberg et al. "Gn-net: The gauss-newton loss for multi-weather relocalization."

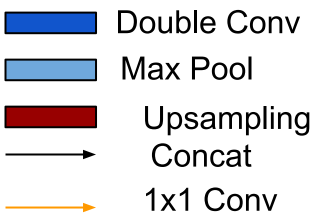
Local Feature Extractor



Input Image
HxW

Output Feature Maps

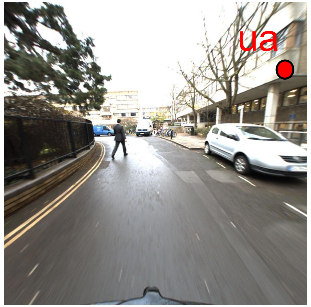
F1	C x H x W
F2	C x H/2 x W/2
F3	C x H/4 x W/4
F4	C x H/8 x W/8

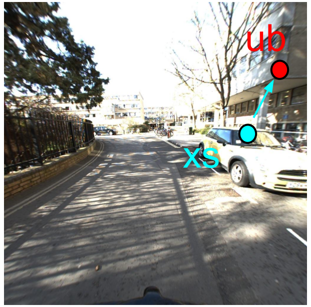


Gauss-Newton Loss

Algorithm 1 Compute Gauss-Newton loss

$\mathbf{F}_a \leftarrow \text{network}(\mathbf{I}_a)$
$\mathbf{F}_b \leftarrow \text{network}(\mathbf{I}_b)$
$e \leftarrow 0$ ▷ Total error
for all correspondences $\mathbf{u}_a, \mathbf{u}_b$ **do**
  $\mathbf{f}_t \leftarrow \mathbf{F}_a(\mathbf{u}_a)$ ▷ Target feature
  $\mathbf{x}_s \leftarrow \mathbf{u}_b + \text{rand}(\text{vicinity})$ ▷ Compute start point
  $\mathbf{f}_s \leftarrow \mathbf{F}_b(\mathbf{x}_s)$
  $\mathbf{r} \leftarrow \mathbf{f}_s - \mathbf{f}_t$ ▷ Residual
  $\mathbf{J} \leftarrow \frac{d\mathbf{F}_b}{d\mathbf{x}_s}$ ▷ Numerical derivative
  $\mathbf{H} \leftarrow \mathbf{J}^T\mathbf{J} + \epsilon \cdot \text{Id}$ ▷ Added epsilon for invertibility
  $\mathbf{b} \leftarrow \mathbf{J}^T r$
  $\boldsymbol{\mu} \leftarrow \mathbf{x}_s - \mathbf{H}^{-1}\mathbf{b}$
  $e_1 \leftarrow \frac{1}{2}(\mathbf{u}_b - \boldsymbol{\mu})^T\mathbf{H}(\mathbf{u}_b - \boldsymbol{\mu})$ ▷ First error term
  $e_2 \leftarrow \log(2\pi) - \frac{1}{2}\log(|\mathbf{H}|)$ ▷ Second error term
  $e \leftarrow e + e_1 + e_2$
end for





Challenges & Solutions

Contrastive loss for negative pairs does not decrease

Sampling negative matches: Progressive mining strategy[3] with distance constraint[4]

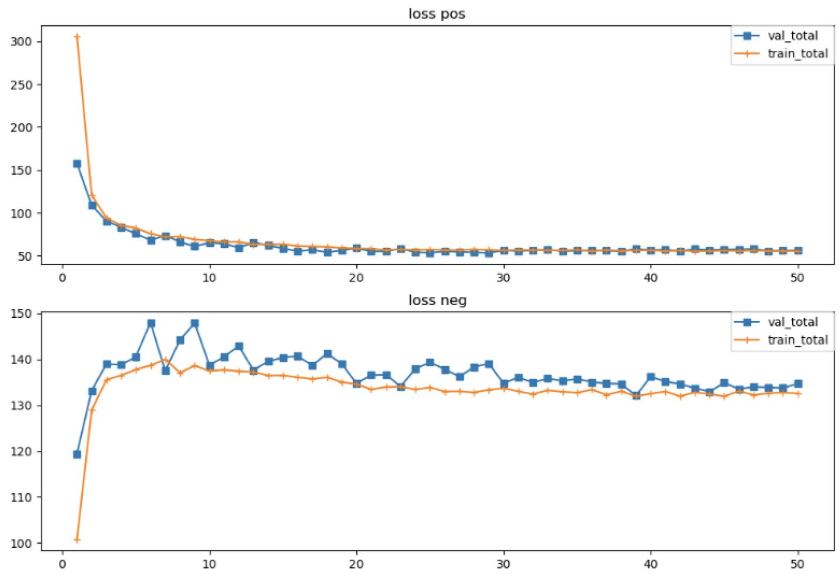
1. Filter out negatives close to positive matches in F_b
2. Sort negatives by loss in increasing order
3. Sample randomly over the smallest top M.

$$M = \max\left(5, 300e^{\frac{-0.6k}{10000}}\right)$$

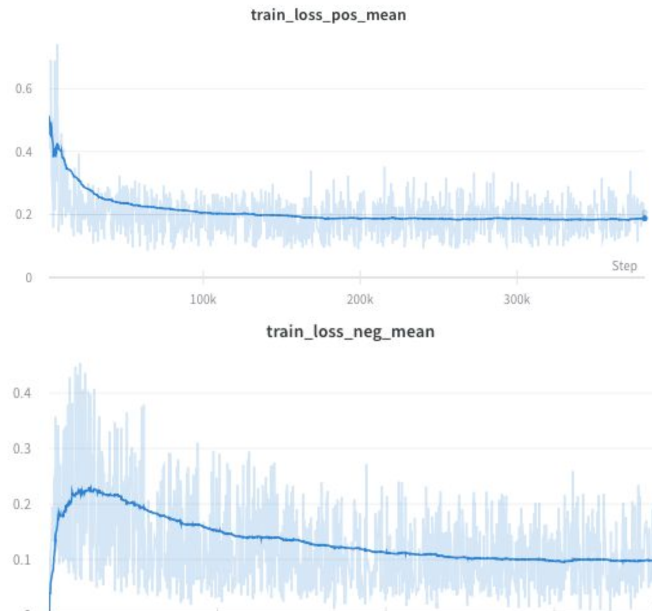
[3] *Yuki et al. "LF-Net: learning local features from images."*
[4] *Dusmanu et al. "D2-net: A trainable cnn for joint detection and description of local features."*

Challenges & Solutions

Contrastive loss for negative matches does not decrease



Randomly sample negatives

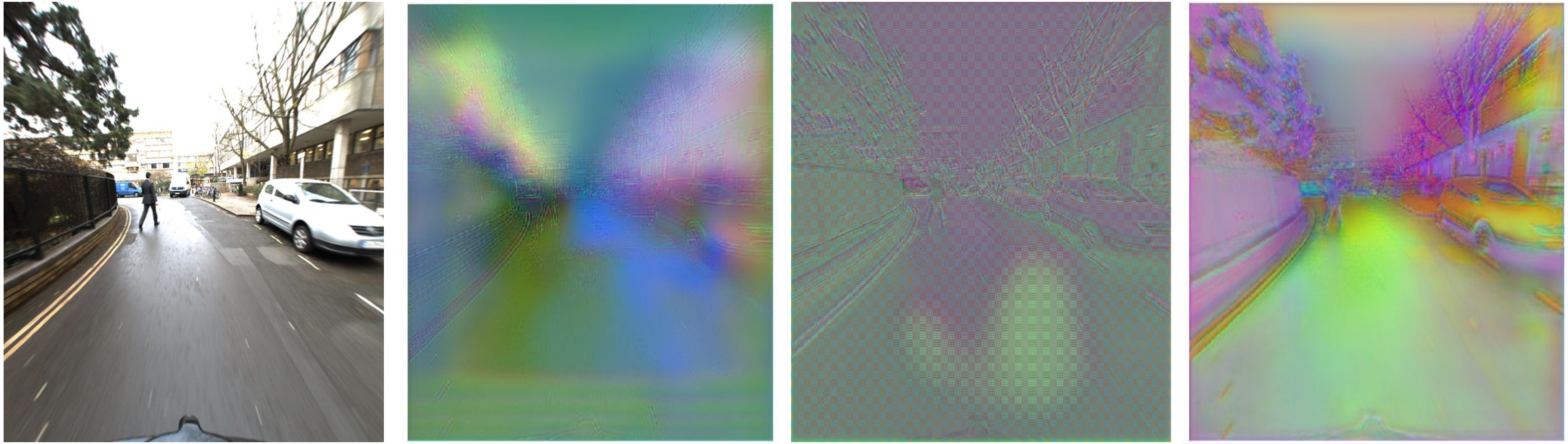


Progressively sample hard negatives

Challenges & Solutions

Contrastive loss not decreasing

Use nearest-neighbor upsampling instead of bilinear interpolation



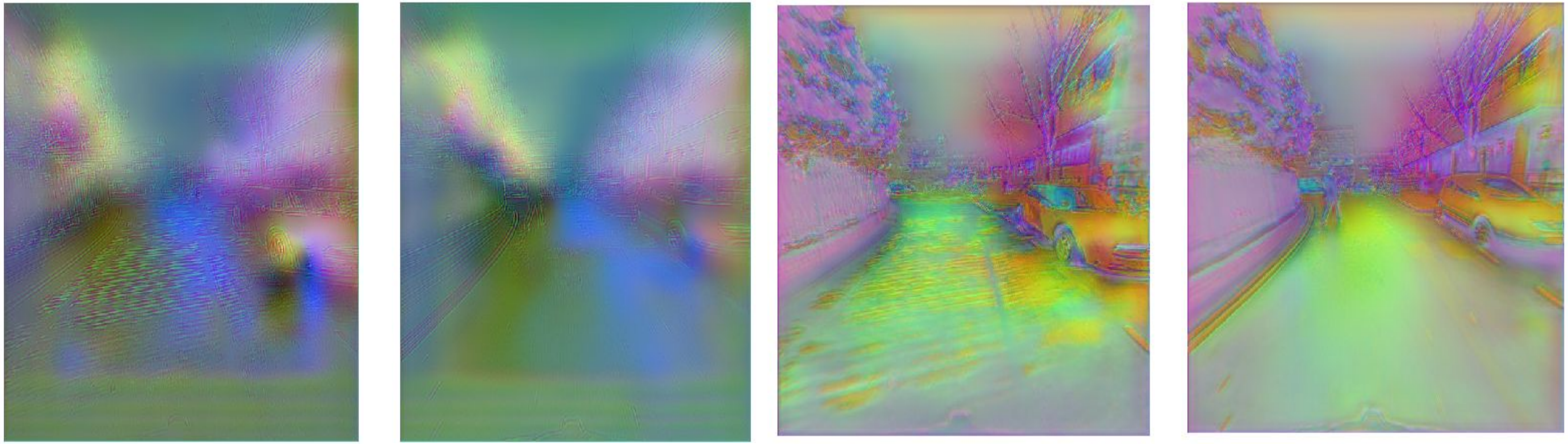
Input | Bilinear | Transpose Conv | Nearest-Neighbor

[5] Odena et al. "Deconvolution and Checkerboard Artifacts."

Challenges & Solutions

Contrastive loss not decreasing

Use nearest-neighbor upsampling instead of bilinear interpolation



Reference Query

Bilinear

Reference Query

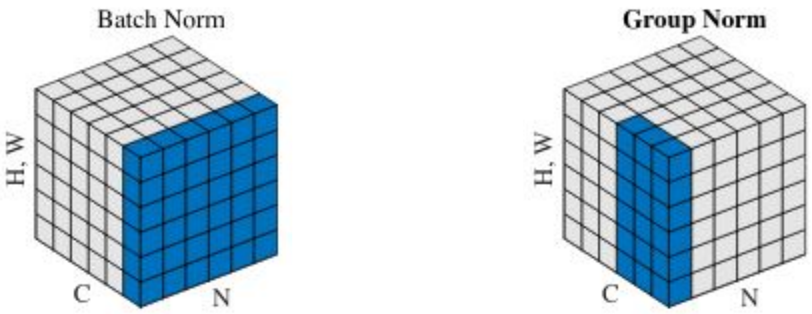
Nearest-Neighbor

Challenges & Solutions

Memory usage increased due to the use of GN-Net and F-PnP

Have to reduce the batch size, feature map resolution, channels etc.

Small batch size leads to inaccurate batch statistics estimation: use Group Normalization[4]



[5] Wu, Yuxin, and Kaiming He. "Group normalization."

Challenges & Solutions

The second term of Gauss-Newton loss does not decrease

Single margin contrastive loss:

$$\mathcal{L}_{\text{contrastive}}\left(\mathbf{F}_a, \mathbf{F}_b, l\right) = \mathcal{L}_{\text{pos}}\left(\mathbf{F}_a, \mathbf{F}_b, l\right) + \mathcal{L}_{\text{neg}}\left(\mathbf{F}_a, \mathbf{F}_b, l\right)$$

$$\mathcal{L}_{\text{pos}}(\mathbf{F}_a, \mathbf{F}_b, l) = \frac{1}{N_{\text{pos}}} \sum_{N_{\text{pos}}} D_{\text{feat}}^2 \qquad \mathcal{L}_{\text{neg}}(\mathbf{F}_a, \mathbf{F}_b, l) = \frac{1}{N_{\text{neg}}} \sum_{N_{\text{neg}}} \max(0, M - D_{\text{feat}})^2$$

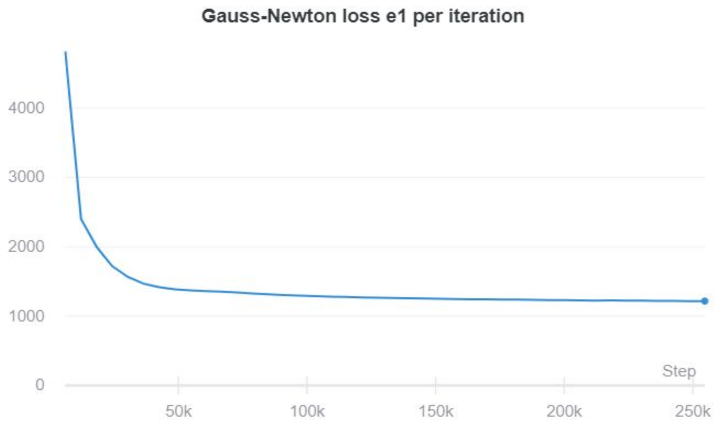
Double margin contrastive loss:

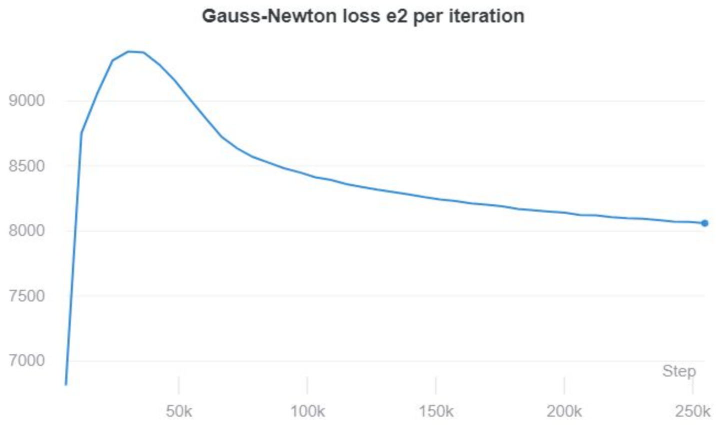
$$\mathcal{L}_{\text{contrastive}}\left(\mathbf{F}_a, \mathbf{F}_b, l\right) = \mathcal{L}_{\text{pos}}\left(\mathbf{F}_a, \mathbf{F}_b, l\right) + \mathcal{L}_{\text{neg}}\left(\mathbf{F}_a, \mathbf{F}_b, l\right)$$

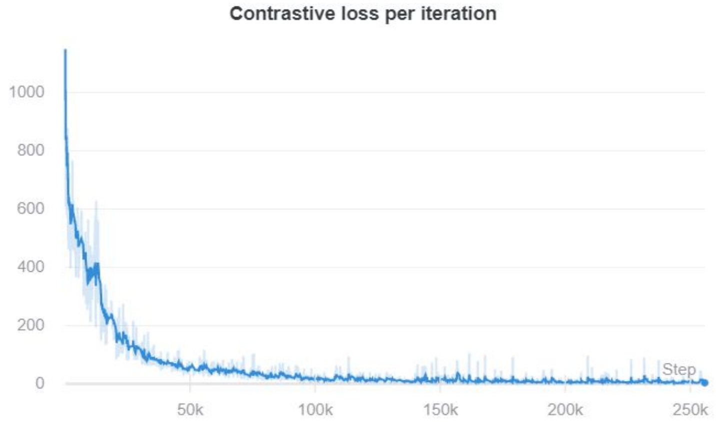
$$\mathcal{L}_{\text{pos}}\left(\mathbf{F}_a, \mathbf{F}_b, l\right) = \frac{1}{N_{\text{pos}}} \sum\nolimits_{N_{\text{pos}}} \max\left(0, D_{\text{pos}} - M_{pos}\right)^2$$

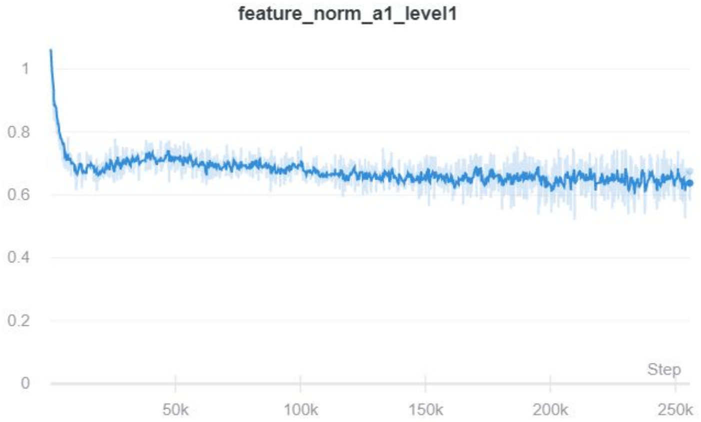
$$\mathcal{L}_{\text{neg}}\left(\mathbf{F}_a, \mathbf{F}_b, l\right) = \frac{1}{N_{\text{neg}}} \sum\nolimits_{N_{\text{neg}}} \max\left(0, M_{neg} - D_{\text{feat}}\right)^2$$

Double margin contrastive loss



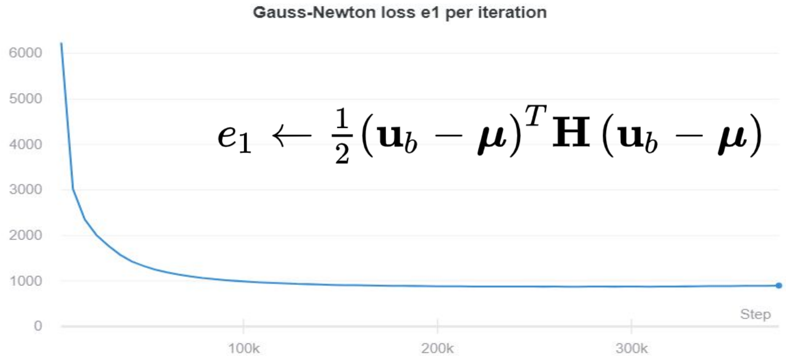




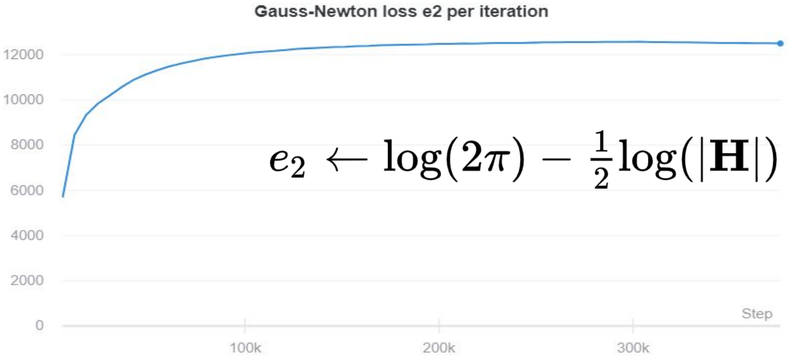


Challenges & Solutions

The second term of Gauss-Newton loss does not decrease

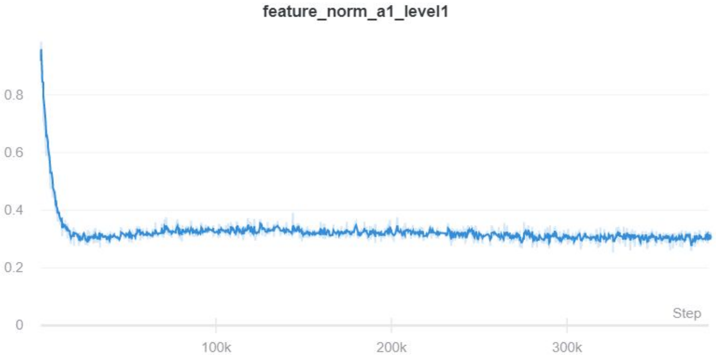


$$e_1 \leftarrow \frac{1}{2}\left(\mathbf{u}_b - \boldsymbol{\mu}\right)^T \mathbf{H} \left(\mathbf{u}_b - \boldsymbol{\mu}\right)$$



$$e_2 \leftarrow \log(2\pi) - \frac{1}{2}\log(|\mathbf{H}|)$$

Possible reasons:
Trained feature norm being too small
-> Hessian with small values.



Challenges & Solutions

Large scale datasets

- Training set for GN-Net [6]:

  28766 pairs for Extended CMU seasons and 6511 pairs for Robotcar:

  Choose a subset of 681 images for fast iteration

- Validation [7]:

  56,613 query images for Extended CMU seasons and 11,934 query images for Robotcar

  Use correspondence and camera intrinsics to estimate relative pose between two images

  Ground truth poses for certain slices of Extended CMU seasons

[6] Larsson et al. "A Cross-Season Correspondence Dataset for Robust Semantic Segmentation."
[7] Sattler et al. "Benchmarking 6DOF Outdoor Visual Localization in Changing Conditions."

Last step

1. Train on the entire Extended CMU Seasons once the cluster is available

2. Validate on slices of CMU with ground truth poses

Thank you!