Foundation models for syntetic aperture radar data: a review and outlook

Stian Normann Anfinsen, Jakob Grahn, Daniel Johansen Trosten, Sigurd Løkse & Filippo Bianchi

NORCE Norwegian Research Centre

ESA-NASA FM4EO workshop, Frascati, Italy, 6 May 2025

Our focus

- How is SAR data used in FMs?
- Do SAR FMs use modality-specific
  - preprocessing?
  - augmentation?
  - pretraining?
- Should they be customised?
- Method: literature review
- Outcome: overview and recommendations for future work

“(…) satellite imagery is a distinct modality and would therefore benefit from modality-specific frameworks and self-supervised learning tasks.”

Lane & Karimzadeh, *arXiv:2504.17177* (2025)

“[one of] the fundamental challenges that need to be addressed (…) [is] the ignorance of SAR characteristics in model design (…)”

Li et al., IEEE TIP, 34:869-884 (2025)

“(…) satellite imagery is a distinct modality and would therefore benefit from modality-specific frameworks and self-supervised learning tasks.”

Lane & Karimzadeh, *arXiv:2504.17177* (2025)

“[one of] the fundamental challenges that need to be addressed (…) [is] the ignorance of SAR characteristics in model design (…)”

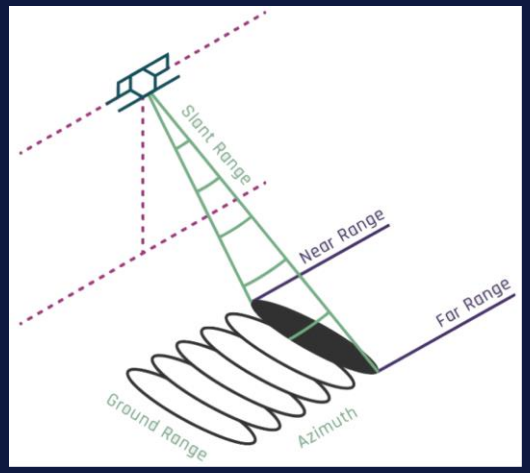
Li et al., IEEE TIP, 34:869-884 (2025)

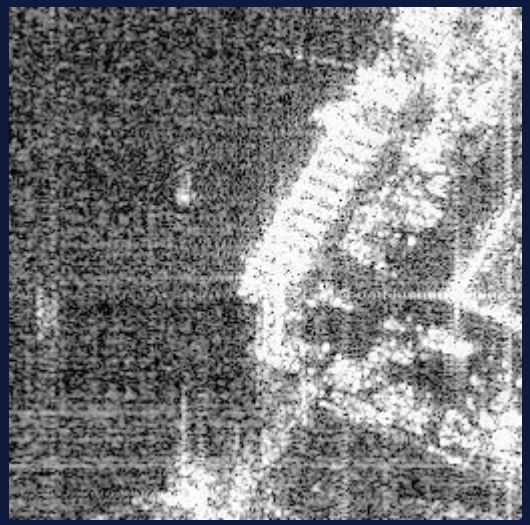
“Traditional models have been (…) tailored to specific sensors or data types (…). This specialization hinders (…) a holistic analysis that could benefit from the combined strength of these diverse data sources ”

Xiong et al., *arXiv:2403.15356* (2024)

SAR characteristics

- Strong, multiplicative noise (speckle)
- Geometric distortion and artefacts (side-looking) sensor)
- Advanced signal processing (frequency domain)
- Special data formats:
  - Single-look complex, Interferometric SAR (InSAR), Polarimetric SAR (PolSAR), Wave mode (WV), etc.





The dilemma: generic or bespoke

- Multimodal or modality-specific?
- How important is it to be modality-aware?
- New augmentation methods needed?
- Adapt pretext to downstream tasks?
- Can pretext tasks be in conflict?

Review

Review includes

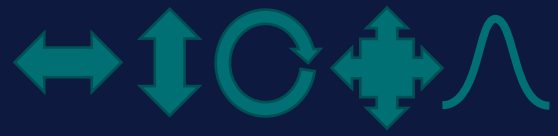
- 50+ papers on:
  - unimodal and multimodal FMs
  - original and survey papers
  - complete and fine-tuned FMs
  - representation learning for SAR
- … and a few papers on fundamental machine learning

Findings in short

- Multimodal FMs do not, in general, customise for SAR data
- Dedicated SAR FMs sometimes do:
  - 2 papers suggest new augmentation
  - 4 papers suggest new pretext tasks
  - 3 models do SAR-specific pretraining
- Most customization is found in FMs for target recognition and PolSAR data

SAR-specific augmentation

- WV-Net for oceanographic Sentinel-1 data tested *random notch filtering of wave spectrum* components in CL approach (Glaser et al., arXiv, 2024)
- (Li et al., IEEE TIP, 2025) reviewed several augmentations for target recognition when designing the SARATR-X model for target recognition (vessel rotation prediction; added noise; feature descriptors)
- (Gallego-Mejia et al., arXiv, 2023) avoided certain augmentations for SAR



SAR-specific pretext tasks

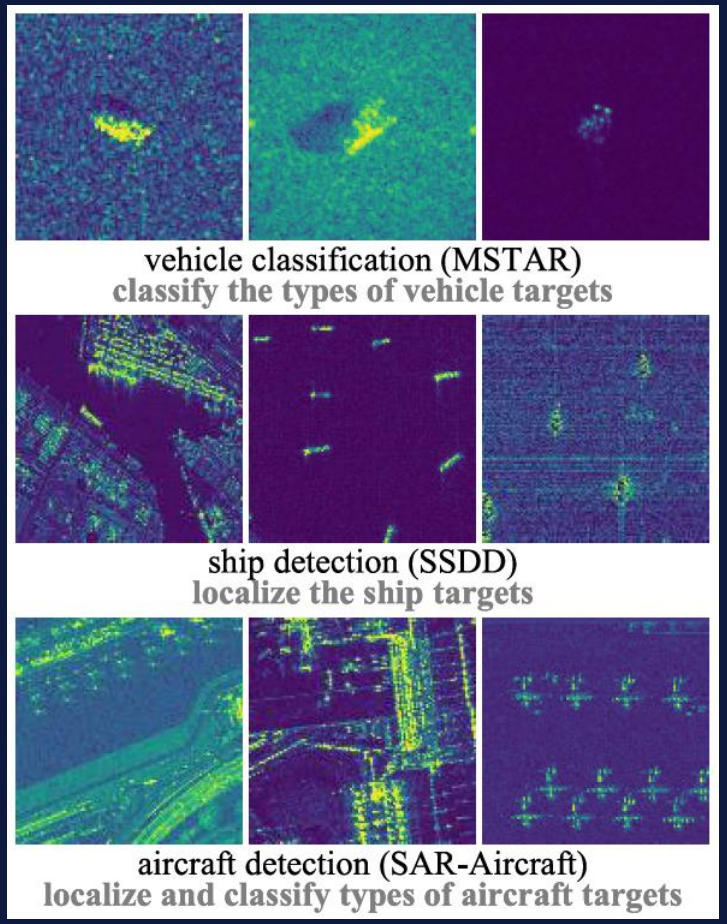
- For complex-valued polarimetric SAR data:
  - The RingMoE model reconstructs the power of polarimetric channels (Bi et al., *arXiv:2504.03166*, 2025)
  - The PolSAM model (Y. Wang et al., *arXiv:2412.12737v1*, 2024) and the model of (M. Wang et al., *arXiv:2504.11999v1*, 2025) reconstructs the scattering power of polarimetric decompositions

SAR-specific pretext tasks

- In automatic target recognition with SAR data:
  - RotANet predicted the rotational patterns of vehicles (Wen et al., IEEE TGRS, 2021)
- In SAR image classification:
  - PGIL used CL between complex SAR image sub-frequency feature and deep amplitude image features to learn physical information (Huang et al., ISPRS JPRS, 2022)

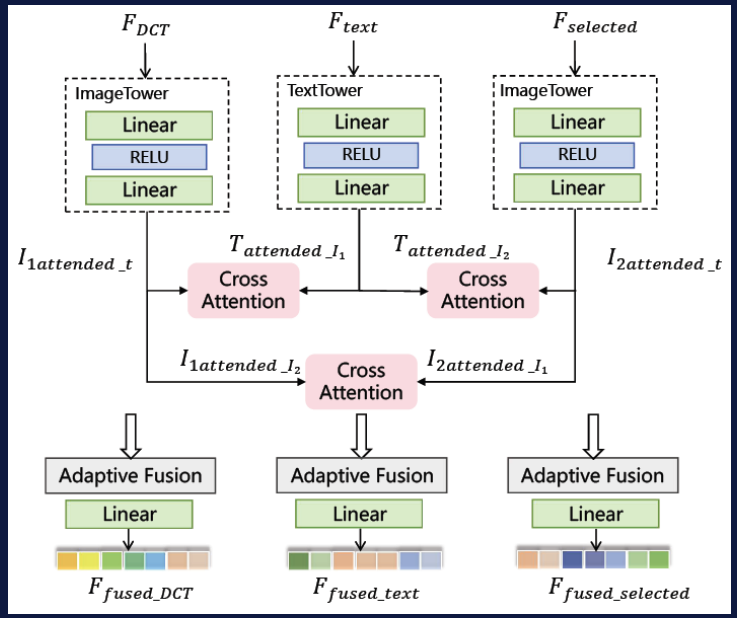
SAR-specific pretraining

- SARATR-X (Li et al., IEEE TGRS, 2025)
- MAE-based FM for automatic target recognition
- Reviews ATR approaches and SSL for SAR:
  - use frequency information, multistage training from ImageNet to SAR, and feature descriptors to suppress noise and enhance the target
- Chosen methodology:
  - Initialises MAE with natural image weights
  - Uses multiscale gradient features to suppress speckle



SAR-specific pretraining

- VLF-SAR (Xie et al., IEEE TCSVT, 2025)
- Vision-language framework for few-shot target recognition
- Reviews several papers that integrate frequency domain information to capture robust and generalised features
- Uses dedicated modules to extract frequency information and polarimetric features



Outlook

What is missing?

- Models cover:
  - amplitude data
  - coherence magnitude data
  - multilook complex PolSAR data
  - Geocoded data
- Not:
  - single-look complex data
  - complex phase data
  - data in SAR sensor geometry

Frequency information

- Proven significant in target recognition
- Contains important physical information for man-made targets
- Frequency-aware loss functions exist
- Can more general and elegant pretext tasks and pretraining frameworks be designed – with benefits beyond target detection?

Speckle

- SAR-W-MixMAE model uses pixel-wise weighting during pretraining to combat speckle (Caglayan et al., *arXiv:2503.01181v2*, 2025)
- Implicitly mitigated by SAR-specific approaches in target recognition models
- However – the despeckling is used as pretext for efficient pretraining of the MERLIN-Seg model (Dalsasso et al., CVIU, 2024)
- Should be used to learn SAR noise characteristics

Temporal dimension

- Sentinel-1 provides dense time series, but speckle is an issue
- No SAR-specific work on SSL to enforce time-variant and time-invariant embeddings, e.g. SAR versions of SeCo, SAR-Caco, SatMAE) and support work with TS
- Extension to spatio-temporal models is important (Li et al., SeaMo, *arXiv:2412.19237v2*, 2025)
- Conflicting pretext tasks can be handled (see e.g. Xiao et al., *arXiv:2008.05659v2*, 2021)

Complex-valued data

- FMs for single-look complex (SLC) data and complex InSAR coherence are needed
- A multitask ViT for detection, location and interpretation of deformation from SAR interferograms is a good starting point for InSAR models, but does not use SSL (Abdallah et al., IJAEOG, 2024)
- Need SSL frameworks for SLC data and InSAR phase

Conclusions

- FMs for SAR data are in their infancy: mostly treated as generic images
- Target recognition models show the effect of modality-specific approaches
- Nontrivial SAR formats are not covered
- Methods from DL for SAR should be adopted and extended
- The goal is multipurpose models that jointly embed multiple (in)variances, using diverse and potentially conflicting pretext tasks, by exploiting methods such as mixture of experts and separate embedding spaces

Thank you. Takk. Merci. Gracias. Obrigado.

Stian Normann Anfinsen
stia@norceresearch.no

norceresearch.no

NORCE