

## Original Research Paper

# Enhancing SAR Image Registration: A Light Weight Deep Network and a Novel Feature Fusion Method

Mohammad Sadegh Helfroush<sup>1\*</sup>  and Sahar Abdi<sup>2</sup>

1. Department of Electrical Engineering, Shiraz University of Technology, Shiraz, Iran
2. Institute of Mechanics, Iranian Space Research Center (ISRC), Shiraz, Iran

**ARTICLE INFO****Article History:**

Received 13 September 2025

Revised 18 October 2025

Accepted 20 October 2025

Available Online 03 November 2025

**Keywords:**

Synthetic aperture radar (SAR)  
Deep learning  
Feature fusion  
Keypoint description  
Computer vision  
Image registration

**ABSTRACT**

Synthetic aperture radar (SAR) image registration is a crucial preprocessing step for many remote sensing applications. However, the presence of speckle noise and homogeneous textures in SAR images makes this task challenging. In this study, we propose a method for SAR image registration. The proposed method, SARNet, is a lightweight and optimized deep neural network specifically designed for SAR images. It generates three distinct descriptors for each keypoint to increase the accuracy of keypoint matching. The method was evaluated on a diverse dataset including Radarsat, Sentinel-1, ALOS-PALSAR, and ERS-2 images. Experimental results show that (SARNet), along with a sub-descriptor matching algorithm, significantly establishes more accurate keypoint correspondences. The network achieves comparable accuracy to state-of-the-art deep learning methods while having the lowest number of parameters (115K) among the methods compared.

\* Corresponding Author's E-mail: [ms\\_helfroush@sutech.ac.ir](mailto:ms_helfroush@sutech.ac.ir)

**How to Cite this Article:**

M. S. Helfroush and S. Abdi, "Enhancing SAR image registration: A light weight deep network and a novel feature fusion method," *Journal of Space Science and Technology*, Vol. ??, No. ?, pp. 1-16, 2026, (in Persian), <https://doi.org/10.22034/jsst.2025.1595>.

**COPYRIGHTS**

© 2026 by the authors. Published by ARI. This article is an open access article distributed under the terms and conditions of [The Creative Commons Attribution 4.0 International \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)



# بهبود تثبیت تصاویر رادار دهانه مصنوعی مبتنی بر یک شبکه عمیق سبک وزن و روش نوین تلفیق ویژگی

محمدصادق هل فروش<sup>۱\*</sup> و سحر عبدی<sup>۲</sup>

۱- استاد، دانشکده مهندسی برق و الکترونیک، دانشگاه صنعتی شیراز، شیراز، ایران

۲- دکترا، پژوهشکده مکانیک، پژوهشگاه فضایی ایران، شیراز، ایران

## چکیده

## اطلاعات مقاله

تثبیت تصاویر رادار دهانه مصنوعی (SAR) یکی از پیش‌پردازش‌های حیاتی برای بسیاری از کاربردهای سنجش از دور است. با این حال، وجود نویز لکه‌ای و بافت‌های همگن در تصاویر SAR این کار را با چالش همراه می‌کند. در این پژوهش روشی برای تثبیت تصاویر SAR پیشنهاد می‌شود. روش پیشنهادی (SARNet)، یک شبکه عصبی عمیق سبک و بهینه مختص تصاویر SAR است که برای هر نقطه کلیدی سه توصیف‌گر مجزا تولید می‌کند تا دقت ایجاد تناظر میان نقاط ویژه را افزایش دهد. این روش بر روی مجموعه داده‌ای متنوع شامل تصاویر Sentinel-1، Radarsat و ALOS-PALSAR و ERS-2 ارزیابی شده است. نتایج تجربی نشان می‌دهد که SARNet به همراه الگوریتم تناظر زیرتوصیف‌گری به طور قابل توجهی تناظرات دقیق‌تری میان نقاط ویژه برقرار می‌کند. این شبکه دقتی قابل قیاس نسبت به روش‌های یادگیری عمیق پیشرفته دارد، در حالی که کمترین تعداد پارامتر (۱۱۵ هزار) را در بین روش‌های مقایسه شده داراست.

## تاریخچه مقاله:

دریافت ۲۲ شهریور ۱۴۰۴

بازنگری ۲۶ مهر ۱۴۰۴

پذیرش ۲۸ مهر ۱۴۰۴

اولین انتشار ۱۲ آبان ۱۴۰۴

## واژه‌های کلیدی:

رادار دهانه مصنوعی

یادگیری عمیق

تلفیق ویژگی

توصیف نقاط ویژه

بینایی ماشین

تثبیت تصاویر

پست الکترونیکی نویسنده مسئول: [ms\\_helfroush@sutech.ac.ir](mailto:ms_helfroush@sutech.ac.ir)

## How to Cite this Article:

M. S. Helfroush and S. Abdi, "Enhancing SAR image registration: A light weight deep network and a novel feature fusion method," *Journal of Space Science and Technology*, Vol. ??, No. ?, pp. 1-16, 2026, (in Persian), <https://doi.org/10.22034/jsst.2025.1595>.



## COPYRIGHTS

© 2026 by the authors. Published by ARI. This article is an open access article distributed under the terms and conditions of [The Creative Commons Attribution 4.0 International \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)



## علائم و اختصارات

$I$	تصور ورودی به شبکه
$S$	خروجی نقشه حرارتی تکرارپذیری
$X$	خروجی توصیف‌گرهای شبکه
$p$	پنجره‌ی هم‌پوشان از نقشه حرارتی تکرارپذیری
$Z$	متغیر تصادفی فاصله اقلیدسی بین توصیف‌گرها
$d$	تک توصیف‌گر
$Cosim$	زاویه کسینوسی بین توصیف‌گرها
$MSE$	میانگین مربع خطا
$AP$	میانگین دقت
$FastAP$	الگوریتم سریع میانگین دقت
$CDF$	تابع توزیع تجمعی
$ \cdot $	قدر مطلق یا تعداد اعضای مجموعه
$N_{att}$	تعداد تمام تناظرات اولیه
$NOCC$	تعداد تناظرات صحیح
$ROCC$	نسبت تناظرات صحیح

## مقدمه

SAR یک سامانه تصویربرداری فعال است که کاربردهای متنوعی در زمینه‌های مختلف از جمله زمین‌شناسی، کشاورزی، تلفیق اطلاعات، تشخیص تغییرات و بسیاری موارد دیگر دارد. مزیت متمایزکننده آن نسبت به دیگر روش‌های تصویربرداری قابلیت تصویربرداری آن در شرایط جوی و نوری مختلف است. به این معنا که می‌توان در تاریکی شب و هوای ابری تصاویری با کیفیت و دقت بالا از زمین تهیه کرد [۱].

پردازش حجم عظیمی از داده‌های این سامانه نیازمند پیش‌پردازش‌هایی است که امکان استفاده از آن‌را در کاربردهای مختلف امکان‌پذیر می‌کند. یکی از این پیش‌پردازش‌های مهم تثبیت تصاویر است. تثبیت به معنای تعریف تصاویر اخذ شده از یک ناحیه در یک دستگاه مختصات واحد است. این تصاویر می‌توانند در زمان‌ها، زوایا و یا حتی توسط حسگرهای مختلفی اخذ شده باشند. تثبیت دو تصویر با یافتن نقاط تکرارپذیر در تصاویر آغاز می‌شود. ویژگی منحصر به فرد این نقاط این است که نسبت به تغییرات زاویه دید، شدت روشنایی، نویز و بسیاری عوامل دیگر مقاوم هستند و مکان آن‌ها تغییر نمی‌کند. با ایجاد تناظر میان نقاط ویژه نواحی مشابه می‌توان تناظر پیکسل به پیکسل میان دو تصویر ایجاد کرد. برای برقراری تناظر میان نقاط ویژه، هر نقطه به کمک اطلاعات

تصویری نواحی اطراف آن توصیف می‌گردد که به آن توصیف‌گر نقاط ویژه گفته می‌شود. توصیف‌گرها بردارهایی هستند که اطلاعات این نواحی را به گونه‌ای کد می‌کنند که به راحتی به توان تناظرات صحیحی میان نقاط ویژه برقرار نمود. ایجاد این تناظرات با استفاده از معیارهایی از جمله فاصله اقلیدسی یا کسینوسی میان توصیف‌گرها انجام می‌پذیرد.

به صورت عمده روش‌های تثبیت تصاویر به دو دسته الگوریتم‌های کلاسیک و روش‌های مبتنی بر یادگیری عمیق تقسیم‌بندی می‌شوند. در هر دسته، مجموعه‌ای از نقاط تکرار پذیر و پایدار در سطح تصاویر یافت و اطلاعات پیرامون آن نقاط به عنوان توصیف‌گر کد می‌شوند. با استفاده از این نقاط و توصیف‌گرهای آن‌ها می‌توان تناظری یک به یک میان نقاط متناظر برقرار کرد که به کمک آن‌ها می‌توان ارتباط ریاضی میان محل پیکسل‌های یک تصویر را با تصویر دیگر مشخص کرد. تفاوت روش‌های کلاسیک و روش‌های مبتنی بر یادگیری عمیق در تشخیص و توصیف این نقاط ویژه نهفته است. در روش‌های کلاسیک نقاط ویژه و توصیف‌گرهای آن‌ها به صورت دست‌ساز امپندسی می‌شوند. ولی در روش‌های مبتنی بر یادگیری عمیق این اطلاعات با توجه به نوع داده‌ها به صورت خودکار و بهینه از تصاویر استخراج می‌گردند.

یکی از مهم‌ترین و پر استفاده‌ترین الگوریتم‌های تثبیت تصاویر SIFT [۲] است. این الگوریتم با استفاده از فضای اختلاف گوسی (DoG) نقاط کلیدی تصویر را مشخص و با استفاده از هیستوگرام گرادینان نواحی اطراف آن نقاط آن‌ها را توصیف می‌کند. توصیف‌گرها نسبت به تغییرات مقیاس، چرخش، تغییر زاویه دید و تا حدودی روشنایی تصاویر مقاوم هستند. با الهام از این الگوریتم، روش‌های متنوعی در طول دو دهه اخیر برای تشخیص و توصیف نقاط ویژه ابداع و معرفی شدند که از آن جمله می‌توان به ASIFT [۳]، BSIFT [۴]، RootSIFT [۵] و بسیاری موارد دیگر اشاره کرد. الگوریتم SURF [۶] با استفاده از انتگرال تصاویر و تبدیل پوولت دقت و کیفیت نقاط ویژه و توصیف‌گرهای استخراجی را بهبود بخشید. الگوریتم KAZE [۷] و نسخه سریع‌تر آن AKAZE [۸] با استفاده از ماتریس Hessian و فضای مقیاس غیر خطی، نقاط ویژه و توصیف‌گرهای مقاوم‌تری ایجاد کردند. برای دستگاه‌های کم‌مصرف، ORB [۹] و BRIEF [۱۰] توصیف‌گرهایی باینری استخراج می‌کنند که مقایسه بین آن‌ها سریع‌تر و بهتر از سایر روش‌ها است. گرچه این افزایش سرعت در اجرا و کاهش بار محاسباتی تنها با کاهش دقت توصیف‌گرها ممکن شده است. یکی از مهم‌ترین چالش‌های تثبیت تصاویر در سال‌های اخیر، تثبیت تصاویر در حضور تغییرات شدید بافت و نوع تصاویر یا تغییر حسگر تصویربرداری است. به عنوان مثال می‌توان به تثبیت تصاویر SAR و نوری به عنوان یکی از چندین نمونه این دسته از تصاویر اشاره کرد. الگوریتم‌های RIFT-2 [۱۱]

و POS-GIFT [۱۲] از جمله مواردی هستند که سعی بر حل مسئله تثبیت تصاویر در مدهای مختلف کرده‌اند.

تصاویر SAR دارای ویژگی‌های خاصی از جمله نویز ضرب‌شونده و بافت بسیار مشابه در سطح تصویر هستند که پردازش آن‌ها را برای تثبیت تصاویر دشوار می‌کند. لذا با در نظر گرفتن این ویژگی‌های خاص الگوریتم‌های اختصاصی برای تصاویر SAR ایجاد و توسعه داده شده‌اند. الگوریتم SAR-SIFT [۱۳] با استفاده از تعریف گرادیانی که نسبت به نویز ضرب‌شونده حساسیت کمتری دارد، دقت تشخیص و توصیف نقاط ویژه را بهبود بخشید. الگوریتم PSO-SIFT [۱۴] روش متفاوتی برای محاسبه گرادیان برای تصاویر SAR معرفی کرد که آن را نسبت به تغییرات غیرخطی شدت روشنایی مقاوم‌تر می‌کند. در KAZE-SAR [۱۵] نقاط ویژه با استفاده از الگوریتم KAZE و توصیف‌گرها با استفاده از الگوریتم SURF برای تثبیت تصاویر SAR استخراج می‌شوند که به مراتب دقت و عملکردی بهتر از هر دوی این الگوریتم‌ها به‌صورت مجزا بر روی تصاویر SAR از خود نشان می‌دهند. مرجع [۱۶] با استفاده از تجانس فاز، توصیف‌گرها و نقاط ویژه معتبرتری ایجاد کردند که حتی در برابر تغییر حسگر تصویربرداری نیز مقاوم است.

روش‌های مبتنی بر یادگیری عمیق در چند سال اخیر باعث پیشرفت‌های چشم‌گیری در تمام حوزه‌های علمی شده است. پردازش تصاویر با استفاده از شبکه‌های عصبی عمیق باعث افزایش دقت و کارایی در کاربردهای مختلفی از جمله تثبیت تصاویر شده است. روش TILDE [۱۷] یکی از نخستین روش‌های مبتنی بر یادگیری عمیق برای تثبیت تصاویر است. در این روش توصیف‌گرها و نقاط ویژه الگوریتم SIFT با استفاده از یک شبکه عمیق تخمین زده شده‌اند. شبکه D2Net [۱۸] نقاط ویژه تکرار پذیر و توصیف‌گرهای با کیفیت در حضور تغییرات شدید شدت روشنایی از تصاویر استخراج می‌کند. شبکه R2D2 [۱۹] قابلیت تشخیص نواحی مناسب برای استخراج نقاط ویژه را دارد. نواحی با بافت مشابه و الگوهای تکراری برای استخراج نقاط ویژه نامناسب هستند و ایجاد تناظر را دشوار می‌کنند. شبکه ALIKED [۲۰] تشخیص نقاط ویژه را به‌صورت یک تابع بهینه مشتق‌پذیر باز تعریف کرد که منجر به افزایش تکرارپذیری نقاط ویژه استخراجی شد. رویکرد DeDoDe [۲۱] استخراج و توصیف نقاط ویژه را از یکدیگر مجزا کرده و هر کدام را به‌صورت مجزا انجام می‌دهد. با این کار هر بخش را می‌توان به‌صورت مجزا بهبود و توسعه داد. گرچه رویکردهای یاد شده به‌صورت ویژه برای تثبیت تصاویر SAR ایجاد نشده‌اند ولی با این وجود کارایی قابل قبولی برای تثبیت تصاویر SAR از خود نشان داده‌اند.

تعداد پژوهش‌های انجام شده برای تثبیت تصاویر SAR، با استفاده از روش‌های یادگیری عمیق، محدود تر از تصاویر نوری هستند. از آن جمله می‌توان به RotNet [۲۲] برای تخمین میزان چرخش دو تصویر نسبت

به یکدیگر نام برد. مرجع [۲۳] شبکه‌ای برای تخمین میزان جابه‌جایی دو تصویر SAR نسبت به یکدیگر معرفی کرده است. شبکه SDNet [۲۴] برای تثبیت تصاویر SAR و نوری معرفی شده است. در این روش قطعه‌هایی از تصاویر نوری و SAR که حول نقاط ویژه خاص و تکرارپذیر استخراج می‌شوند، برای برقراری تناظر میان تصاویر مورد استفاده قرار گرفته‌اند. مرجع [۲۵] برای حذف نقاط ویژه فاقد تناظر در میان تصاویر SAR از روش‌های دسته‌بندی استفاده کرده است. مرجع [۲۶] برای تثبیت تصاویر نوری و SAR فضای ویژگی مشترکی بین تصاویر SAR و نوری ایجاد کرده و با استفاده از این ویژگی‌ها تناظرات معتبری بین تصاویر SAR و نوری ایجاد کرده است. مرجع [۲۷] با استفاده از یک شبکه عصبی عمیق، تبدیل هندسی میان جفت تصاویر SAR را از تصاویر ورودی به‌صورت مستقیم محاسبه کرده است. MFAM-RegNet [۲۸] با استفاده از لبه‌ها در تصاویر SAR و نوری، نقاط ویژه مشترکی تعریف کرده است. این نقاط ویژه راس‌های گرافی را تشکیل می‌دهند. سپس تناظرات یک به یکی میان این راس‌ها برقرار می‌شود.

از دیگر جنبه‌های مهم در تعیین دقت تثبیت تصاویر می‌توان به روش‌های ایجاد تناظر میان نقاط ویژه اشاره کرد. هرچه دقت ایجاد تناظرات بیشتر باشد دقت مدل هندسی نهایی برای انجام تثبیت نیز بیشتر خواهد بود. برای این کار غالباً توصیف‌گر میان نقاط با یکدیگر مقایسه شده و شبیه‌ترین آن‌ها متناظر می‌گردند. رویکرد NN [۲۹] از فاصله اقلیدسی یا همبستگی برای تعیین شبیه‌ترین توصیف‌گرها استفاده می‌کند. روش دیگر انتخاب همسایگی‌های نزدیک و متقابل است که از آن به‌عنوان MNN [۳۰] یاد می‌شود. با این کار خطا به‌شدت کاهش می‌یابد. مرجع [۲] برای کاهش خطا، تناظراتی را صحیح انگاشت که نسبت فاصله نزدیک‌ترین همسایگی تا دومین همسایگی نزدیک از عدد مشخصی بیشتر باشد. با این کار تا حدود زیادی خطای تطبیق نقاط متناظر کاهش می‌یابد. در روش FLANN [۳۱] تخمینی سریع‌تر از روش NN معرفی کرد. مرجع [۳۲] با استفاده از چندین توصیف‌گر و تقسیم آن‌ها به زیربخش‌های کوچکتر تناظرات دقیق‌تری ایجاد کرد.

تناظرات اولیه ایجاد شده شامل خطاهایی هستند که باعث خطا و یا حتی شکست در تثبیت خواهند شد. برای حذف تناظرات اشتباه الگوریتم‌های حذف تناظرات پرت وجود دارند. برخی از این رویکردها هم‌زمان مدل هندسی میان تصاویر را نیز تخمین می‌زنند. از جمله معروف‌ترین آن‌ها می‌توان به روش RANSAC [۳۳] اشاره نمود. با الهام از این الگوریتم روش‌های زیادی معرفی شده‌اند که از آن جمله می‌توان به USAC [۳۴] اشاره نمود. تفاوت عمده این الگوریتم با RANSAC در پایداری بیشتر آن در حضور تعداد تناظرات پرت است. روش‌های مبتنی بر یادگیری نیز برای تعیین تناظرات صحیح مورد استفاده قرار می‌گیرند. از آن جمله می‌توان به SuperGlue [۳۵] و

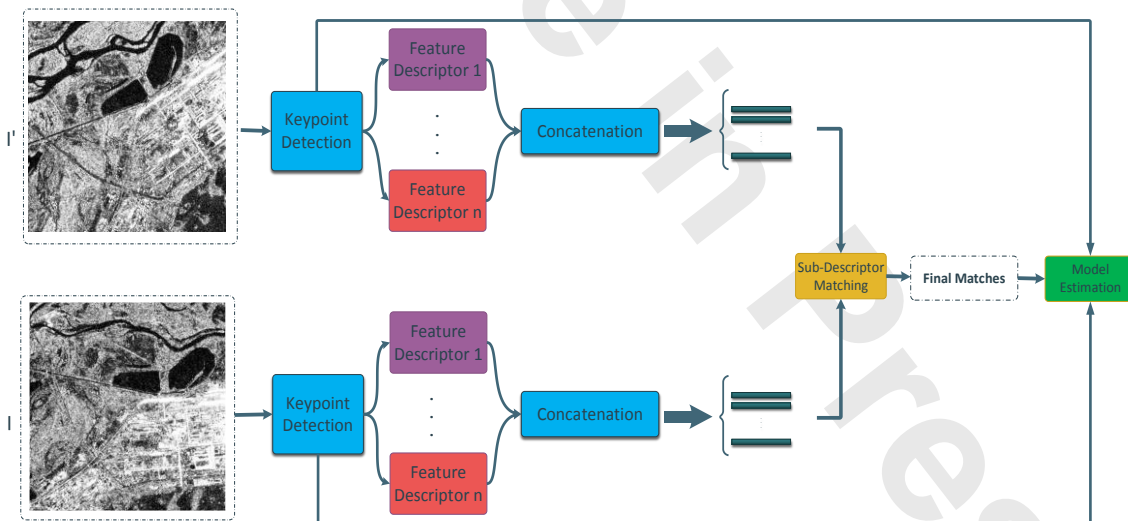
مجزا برای هر نقطه کلیدی به منظور بهبود تطبیق نقاط کلیدی. مابقی این مقاله به شرح زیر سازماندهی شده است: در بخش ۲، روش پیشنهادی را معرفی می‌کنیم. بخش ۳ به نتایج شبیه‌سازی و تحلیل‌ها خواهد پرداخت و در نهایت، بخش آخر به جمع‌بندی و بحث‌های نهایی اختصاص یافته است.

## روش پیشنهادی

شکل ۱ نمای کلی از روش پیشنهادی را نشان می‌دهد. این روش با استخراج نقاط کلیدی از تصاویر SAR آغاز می‌شود. برای هر نقطه کلیدی استخراج‌شده، چندین توصیف‌گر متمایز تولید شده و سپس توصیف‌گرهای هر نقطه ویژه با یکدیگر تلفیق می‌گردند. تطابق بین نقاط کلیدی با استفاده از الگوریتم تطبیق زیرتوصیف‌گرها که در مرجع [۳۲] معرفی شده انجام می‌شود و مدل هندسی بین تصاویر با استفاده از این تطابق‌ها محاسبه می‌گردد. این رویکرد هم بر روش‌های سنتی مبتنی بر مهندسی دستی توصیف‌گرها و هم بر روش‌های مبتنی بر یادگیری اعمال قابل اعمال است.

LightGlue [۳۶] اشاره کرد که از شبکه‌های عصبی گرافی برای یافتن تناظرات صحیح استفاده می‌کنند.

همانطور که ذکر شد، تثبیت تصاویر SAR با وجود کاربردهای مهم در زمینه‌های مختلف، به دلیل ویژگی‌های خاص آن دشوار است. روش‌های کلاسیک قابلیت و دقت محدودی در تثبیت این تصاویر دارند. کاربردهای روزافزون تصاویر SAR نیازمند روش‌های دقیق‌تر و سریع‌تر است. در این تحقیق، با الهام از مطالعات انجام شده بر روی تثبیت تصاویر نوری و با در نظر گرفتن محدودیت‌ها و ویژگی‌های خاص تصاویر SAR، روشی برای تثبیت تصاویر SAR معرفی می‌شود. با الهام از تطبیق ویژگی زیرتوصیف‌گر، روشی مبتنی بر یادگیری عمیق با ایجاد سه توصیف‌گر مجزا برای هر نقطه کلیدی به منظور افزایش دقت تطبیق نقاط ویژه ارائه می‌دهیم. در روش پیشنهادی، بار محاسباتی و دقت، فاکتورهای طراحی مهمی هستند که در نظر گرفته شده‌اند. به طور خلاصه، نوآوری‌های روش‌های پیشنهادی عبارتند از: (۱) اعمال تطبیق زیرتوصیف‌گر به طور خاص برای تصاویر SAR. (۲) معرفی یک شبکه عصبی عمیق سبک و تخصصی برای استخراج و توصیف نقاط کلیدی برای تصاویر SAR. (۳) آموزش یک شبکه برای تولید سه توصیف‌گر



شکل ۱- نمای کلی روش پیشنهادی. از تصاویر ورودی چندین توصیف‌گر برای هر نقطه ویژه استخراج می‌گردد. سپس توصیف‌گرها به یکدیگر الحاق شده و با استفاده از رویکرد زیرتوصیف‌گری تناظرات برقرار می‌شوند. تناظرات نهایی برای تخمین مدل هندسی استفاده می‌شوند.

Fig. 1. Overview of the proposed method. Multiple descriptors are extracted from the input images. The extracted descriptors are then concatenated and then matched using sub-descriptor matching approach. The final matches are used for geometric model estimation.

SAR نشان می‌دهد. تصاویر (الف) و (ب) تطابق‌هایی هستند که به طور مستقل با توصیف‌گرهای KAZE و SIFT به دست آمده‌اند. در هر دو تصویر نقاط ویژه یکسان هستند. تصویر (ج) تطابق‌هایی را نشان می‌دهد که بین هر دو نوع توصیف‌گر مشترک‌اند. همانگونه که مشاهده می‌شود

در مرجع [۲۸] پیشنهاد شده که نتایج تطابق ویژگی‌ها بر روی مجموعه‌ای از نقاط کلیدی یکسان با توصیف‌گرهای متفاوت ترکیب شوند تا تطابق‌های قابل اطمینان تری بین نقاط ویژه به دست آید. شکل ۲ مثالی از این روش را با به کارگیری الگوریتم‌های KAZE و SIFT برای تصاویر

### معماری شبکه SARNet

شکل ۳ معماری شبکه پیشنهاد شده برای استخراج نقاط ویژگی و توصیف‌گرها از تصاویر SAR را نشان می‌دهد. برای هر تصویر ورودی، شبکه یک نقشه حرارتی برای مکان‌یابی نقاط کلیدی و سه توصیف‌گر مجزا تولید می‌کند. ماکزیم‌های محلی در نقشه حرارتی محل نقاط ویژه را تعیین می‌کنند و سه توصیف‌گر مجزا برای ایجاد تناظرات دقیق میان نقاط ویژه مورد استفاده قرار می‌گیرند.

در شبکه پیشنهادی، تصاویر ابتدا از بلوک‌های کانولوشنی ResNet عبور می‌کنند. این بلوک‌ها از تابع فعال‌سازی SELU و Batch Normalization استفاده می‌کنند. در پایان هر بلوک، یک لایه Pooling برای کاهش ابعاد فضای ویژگی اعمال می‌شود که ضرایب کاهش نمونه برای هر بلوک به ترتیب ۲، ۲، ۲ و ۴ است. این کار دو هدف را دنبال می‌کند: (۱) کاهش بار محاسباتی و افزایش سرعت اجرا، و (۲) افزایش میدان دید توصیف‌گرها. با این کار برای هر نقطه ویژگی ناحیه‌ای به اندازه  $32 \times 32$  پیکسل توصیف خواهد شد.

در مرحله بعد ویژگی‌های استخراج شده با استفاده از درون‌یابی دوخطی (Bilinear) به اندازه ابعاد تصویر اولیه تبدیل و سپس به یکدیگر الحاق می‌شوند. سپس این ویژگی‌ها به بلوک‌های مجزا برای استخراج مکان نقاط کلیدی و توصیف‌گرها داده می‌شوند.

بلوک *RepBlock* فضای نهان ویژگی را به ۸ کانال و سپس به تدریج به ۱ کانال تبدیل می‌کند تا نقشه حرارتی برای تعیین محل نقاط ویژه تولید شود. بلوک‌های *DesBlock1*، *DesBlock2*، و *DesBlock3* نیز هر کدام توصیف‌گرهایی با عمق ۶۴ برای هر پیکسل از تصویر ورودی تولید می‌کنند که برای ایجاد تناظرات دقیق‌تر مورد استفاده قرار می‌گیرند.

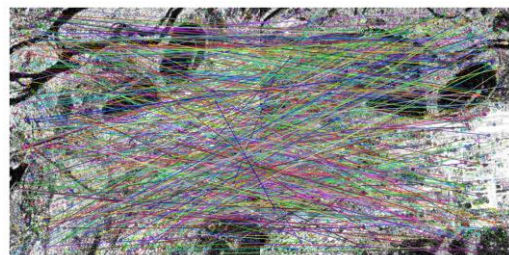
### آموزش شبکه

همانطور که در شکل ۴ نشان داده شده است، فرآیند آموزش شبکه پیشنهادی شامل ورود یک جفت تصویر SAR با یک تبدیل هندسی شناخته شده به شبکه است. خروجی‌های تولید شده توسط شبکه با استفاده از معکوس این تبدیل شناخته شده هم‌مکان می‌شوند. هم‌مکان شدن فضای ویژگی جفت تصاویر ورودی امکان مقایسه فضاهای ویژگی را فراهم می‌کند. خروجی‌های هم‌مکان شده به‌عنوان ورودی برای تابع هزینه عمل می‌کنند و بهینه‌سازی شبکه را هدایت می‌کنند.

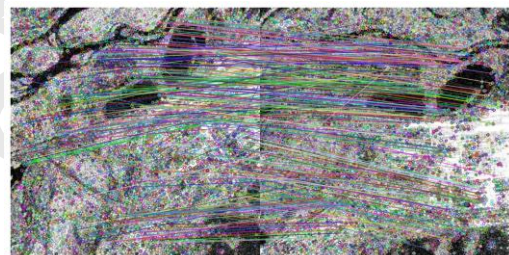
نتایج حاصل از تلفیق تطابق‌ها عموماً قابل اعتمادترند. این رویکرد مبتنی بر ترکیب نتایج برای اظهار نظر نهایی و دقیق‌تر در مورد یک ارزیابی است. برای مثال می‌توان به روش‌های دسته‌بندی تصاویر اشاره کرد. برای بهبود دقت شبکه‌های عصبی مختص دسته بندی تصاویر، یکی از روش‌های مرسوم استفاده از چند شبکه با معماری‌های مختلف برای دسته‌بندی تصاویر است. هر شبکه به‌صورت مجزا در مورد تصاویر اظهار نظر می‌کند و تصمیم‌گیری نهایی بر اساس نظر غالب انجام می‌شود. با الهام از این رویکرد شبکه‌ای عمیق با قابلیت ایجاد توصیف‌گرهای مختلف برای هر نقطه ویژه تصاویر SAR معرفی می‌کنیم که قابلیت ایجاد تناظرات دقیق‌تری را دارد. در ادامه این بخش اجزای مختلف شبکه عمیق پیشنهادی برای تثبیت تصاویر SAR معرفی می‌شوند.



(A)



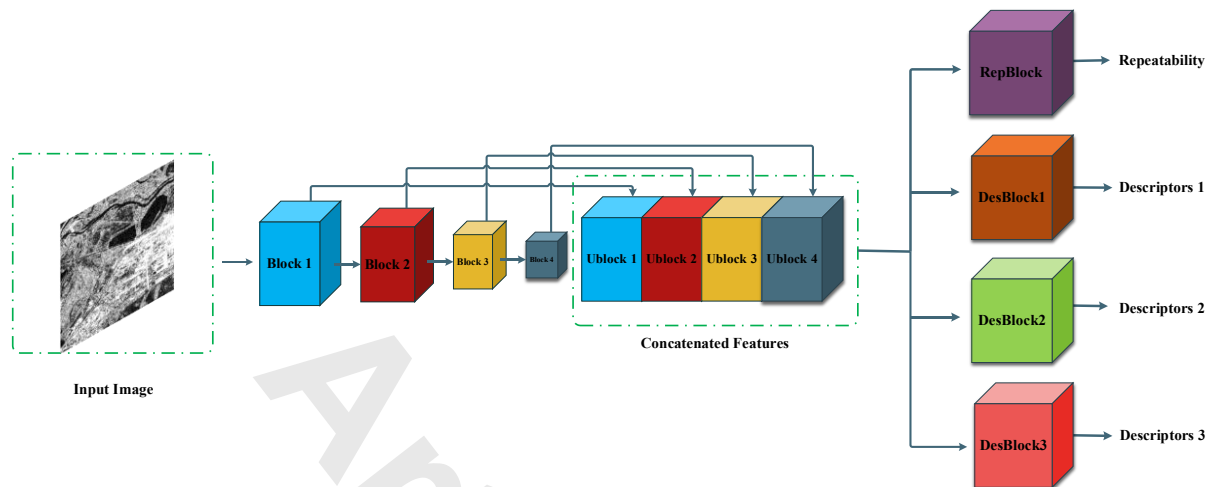
(B)



(C)

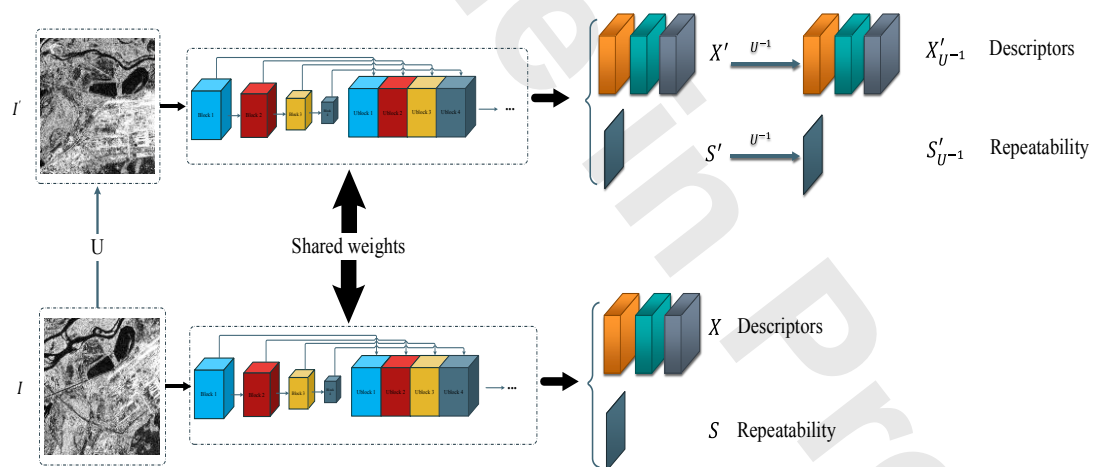
**شکل ۲-** ایجاد تناظر با استفاده از توصیف‌گرهای KAZE و SIFT بر روی تصاویر SAR. نقاط ویژه توسط الگوریتم KAZE مشخص شده‌اند. (الف) تناظرهای برقرار شده پس از توصیف نقاط ویژه با استفاده از KAZE. (ب) تناظرهای برقرار شده پس از توصیف نقاط ویژه با استفاده از الگوریتم SIFT. (د) تناظرهای مشترک (الف) و (ب).

**Fig. 2.** Matches established using KAZE and SIFT descriptors on SAR images for KAZE keypoints. a) Matches for KAZE descriptors. b) Matches for SIFT descriptors. c) Common reoccurring matches between KAZE and SIFT descriptors.



شکل ۳- معماری شبکه پیشنهادی. شبکه‌های پیشنهادی از بلوک‌های کانولوشن ساده‌ای تشکیل شده‌اند که ویژگی‌های مناسب را برای تولید مکان‌های نقاط کلیدی، نواحی قابل اطمینان برای انتخاب نقاط کلیدی و توصیف‌گرها استخراج می‌کنند.

Fig. 3. Proposed network architecture. The proposed networks consist of Simple Convolutional Blocks that extracts appropriate features to output keypoint's locations, reliable areas for keypoint selection and descriptors.



شکل ۴- آموزش شبکه پیشنهادی. یک جفت تصویر با تبدیل هندسی معلوم به شبکه داده می‌شوند. سپس خروجی‌ها با استفاده از مدل تبدیل معلوم هم‌راستا می‌شوند. از خروجی‌های هم‌راستا شده برای آموزش شبکه استفاده می‌شود.

Fig. 4. Training of the proposed network. A pair of images with known transformation are fed to the network. Outputs are then aligned using the known transformation model. The aligned outputs are then used to train the network.

هزینه  $L_{Cosim}$ ،  $L_{Sharpening}$  و  $L_{peaky}$  به کار گرفته می‌شوند. ترکیب این سه تابع تضمین می‌کند که نقاط ویژگی یکسانی در نواحی مشابه از تصاویر استخراج می‌شوند.  $L_{Cosim}$  تضمین می‌کند که نقاط ویژگی در نقشه حرارتی هر دو تصویر در مکان‌های یکسانی قرار می‌گیرند. این تابع، که برای استخراج نقاط ویژگی استفاده می‌شود، به شرح زیر تعریف می‌شود:

### تابع هزینه

آموزش شبکه شامل دو تابع هزینه متمایز برای مکان‌یابی و توصیف نقاط ویژگی است. هدف استخراج نقاط ویژگی، تشخیص نقاط ویژه یکسان در تصاویر مختلف و مکان‌یابی دقیق آن‌ها با استفاده از نقشه‌های حرارتی با ماکزیمم‌های محلی است. برای این منظور سه تابع

که در آن  $S[P]$  نشان‌دهنده پنجره غیرهم‌پوشان از نقشه‌حرارتی تکرارپذیری پیش‌بینی شده در محل  $P$  است.  $\delta[P]$  پنجره معادل  $S[P]$  است که فقط در محل ماکزیمم محلی مقدار غیر صفر یک دارد و  $|P|$  تعداد تمام نقاط ماکزیمم را نشان می‌دهد. با این حال، از نظر کمی، تعداد ماکزیمم‌های محلی به‌طور قابل توجهی کمتر از نقاط غیر ماکزیمم هستند. این عدم تعادل منجر به همگرایی بسیار کند این تابع می‌شود. برای حل این مشکل،  $MSE$  را برای نقاط ماکزیمم محلی و غیر ماکزیمم به‌طور جداگانه به‌صورت زیر محاسبه می‌کنیم:

$$MSE_{maxima} = \frac{1}{|G|} \sum_{(x,y) \in G} |S(x,y) - 1|^2 \quad (5)$$

که در آن  $S(x,y)$  مقدار خروجی نقشه حرارتی تکرارپذیری در محل  $(x,y)$  و  $G$  مجموعه مکان‌های ماکزیمم محلی در نقشه‌حرارتی است و  $MSE_{maxima}$  خطای ماکزیمم‌های محلی در نقشه‌حرارتی خروجی است. به‌طور مشابه می‌توانیم  $MSE_{non-maxima}$  را به‌صورت زیر محاسبه کنیم:

$$MSE_{non-maxima} = \frac{1}{M - |G|} \sum_{(x,y) \notin G} |R(x,y)|^2 \quad (6)$$

که  $M$  تعداد تمام پیکسل‌های موجود در نقشه حرارتی خروجی است. در نهایت  $L_{sharpening}$  به‌صورت زیر محاسبه می‌شود:

$$L_{sharpening} = MSE_{maxima} + \lambda MSE_{non-maxima} \quad (7)$$

که در آن  $\lambda$  ضریب وزن‌دهی است که در اینجا ۴ انتخاب شده است. برای بهینه‌سازی همزمان این توابع، یک تابع هزینه ترکیبی به‌صورت یک مجموع وزن‌دار فرموله می‌شود:

$$L_{rep} = \lambda_1 L_{cosim} + \lambda_2 L_{peaky} + \lambda_3 L_{sharpening} \quad (8)$$

که در آن  $\lambda_1$ ،  $\lambda_2$  و  $\lambda_3$  ضرایب وزن‌دهی هستند که اهمیت نسبی هر جمله را کنترل می‌کنند. این وزن‌ها به‌صورت تجربی تعیین می‌شوند تا اهداف مختلف را متعادل کرده و عملکرد بهینه را به‌دست آورند. در این پژوهش، این مقادیر به ترتیب 0.5، 1.0 و 0.5 تنظیم شده‌اند.

برای آموزش توصیف‌گرها نیازمند معرفی میانگین دقت هستیم. فرض کنید  $X$  فضای ویژگی توصیف‌گر،  $d \in X$  یک توصیف‌گر، و  $R \subset X$  زیرمجموعه‌ای از توصیف‌گرها شامل عناصر متناظر  $R^+$  و غیرمتناظر  $R^-$  با  $d$  باشند. هدف ما نگاشت ویژگی‌های عمیق استخراج‌شده از شبکه به یک فضای برداری  $N$  بعدی است که در آن توصیف‌گرهای متناظر حداقل فاصله اقلیدسی و توصیف‌گرهای غیر متناظر حداکثر

$$L_{Cosim}(I, I', U) = 1 - \frac{1}{|P|} \sum_{p \in P} Cosim(S[p], S'_{U^{-1}}[p]) \quad (1)$$

در اینجا،  $I$  و  $I'$  به ترتیب تصاویر ورودی به شبکه را نشان می‌دهند. رابطه بین تصاویر ورودی با تبدیل هندسی  $U$  تعریف می‌شود. شبکه خروجی‌های نقشه حرارتی  $S$  و  $S'$  را به ترتیب از تصاویر ورودی  $I$  و  $I'$  تولید می‌کند. برای هم‌راستا کردن همه خروجی‌ها، تبدیل هندسی معکوس  $U^{-1}$  به خروجی‌های تصویر تبدیل شده اعمال می‌شود. مقایسه خروجی‌های به‌صورت محلی و پنجره‌های کوچک هم‌پوشان صورت می‌گیرد.  $P$  مجموعه محل پنجره‌های هم‌پوشانی است که از هر دو تصویر با یکدیگر مقایسه می‌شوند و  $p$  محل یکی از این پنجره‌ها را نشان می‌دهد.  $|P|$  تعداد تمام پنجره‌های هم‌پوشان را نشان می‌دهد.  $S[p]$  پنجره‌ای در محل  $p$  از نقشه حرارتی  $S$  را نشان می‌دهد. به‌صورت مشابه  $S'_{U^{-1}}[p]$  پنجره‌ای در محل  $p$  از نقشه حرارتی  $S'_{U^{-1}}$  را نشان می‌دهد. نقشه‌حرارتی تکرارپذیری به بخش‌های همپوشان  $P$  تقسیم می‌شود و شباهت آن‌ها با استفاده از تابع  $Cosim$  محاسبه می‌شود. این تابع، همبستگی یا زاویه‌ی کسینوسی میان تصاویر ورودی را محاسبه می‌کند و به‌صورت زیر تعریف می‌شود.

$$Cosim(a, b) = \frac{a \cdot b}{\|a\| \|b\|} \quad (2)$$

برای جلوگیری از همگرایی نقشه‌حرارتی به یک تصویر ثابت با مقادیر مشابه، از تابع  $L_{peaky}$  استفاده می‌شود. این تابع به‌صورت زیر تعریف می‌شود:

$$L_{peaky}(I) = 1 - \frac{1}{|P|} \sum_{p \in P} \max(S[p]) - \min(S[p]) \quad (3)$$

که در آن  $S[p]$  پنجره‌های کوچکی هم‌پوشاننده از نقشه حرارتی تکرارپذیری  $S$  از تصویر  $I$  هستند.  $\max(\cdot)$  و  $\min(\cdot)$  به ترتیب کمترین و بیشترین مقدار ماتریس ورودی هستند.

هدف  $L_{sharpening}$ ، تیز کردن نقشه‌حرارتی تکرارپذیری با تولید برجسب‌هایی از نقاط اکسترمم محلی است و شبکه را وادار می‌کند تا قله‌های تک پیکسلی بسیار دقیق در این مکان‌ها تولید کند. یک رویکرد ساده برای انجام آن این است که میانگین مربعات خطا  $MSE$  را برای نقشه‌حرارتی تکرارپذیری خروجی و نقشه ماکزیمم‌های محلی تولید شده از نقشه‌حرارتی تکرارپذیری به‌صورت زیر محاسبه کنیم:

$$MSE = \frac{1}{|P|} \sum_{p \in P} |S[p] - \delta[P]|^2 \quad (4)$$

بیان شود، که آن را برای تکنیک‌های بهینه‌سازی مبتنی بر گرادین مناسب می‌سازد. برای تولید توصیف‌گرهای متمایز، ما از FastAP [۳۷] استفاده می‌کنیم.

برای اطمینان از اینکه توصیف‌گرهای مکان‌های یکسان مشابه باشند در حالی که توصیف‌گرهای مکان‌های مختلف متمایز باشند، به صورت شبکه‌ای بر روی توصیف‌گرهای هم‌مکان شده از هر دو تصویر ورودی شبکه نمونه‌برداری انجام می‌شود. سپس هر توصیف‌گر نمونه‌برداری شده با مجموعه‌هایی از توصیف‌گرهای متناظر و نامتناظر از خروجی‌های دیگر مقایسه می‌شود. تابع  $AP$  شباهت بین توصیف‌گرهای متناظر و عدم شباهت بین توصیف‌گرهای نامتناظر را اندازه‌گیری می‌کند. برای استفاده از  $L_{AP}$  به عنوان تابع هزینه، آن را به صورت زیر بازنویسی می‌کنیم:

$$L_{AP}(q) = 1 - AP(q) \quad (۱۶)$$

در اینجا،  $q$  یک توصیف‌گر را نشان می‌دهد که با مجموعه‌هایی از توصیف‌گرهای متناظر و نامتناظر مقایسه می‌شود. در اینجا توصیف‌گر متناظر از خروجی توصیف‌گر یکسان انتخاب می‌شوند ولی توصیف‌گرهای غیر متناظر می‌توانند از تمام خروجی‌ها استخراج گردند. در نهایت، تابع هزینه کلی بر روی تمام مجموعه‌های  $q$  با استفاده از معادله زیر تعریف می‌شود:

$$L_{AP} = \frac{1}{|Q|} \sum_{q \in Q} L_{AP}(q) \quad (۱۷)$$

که در آن که در آن  $|Q|$  تعداد مجموعه‌های نمونه‌برداری شده است. سپس تابع هزینه کل به صورت زیر تعریف می‌شود:

$$L_{total} = L_{AP} + \lambda L_{Rep} \quad (۱۸)$$

که در آن  $L_{AP}$  تولید توصیف‌گرهای متمایز را کنترل می‌کند و  $L_{Rep}$  تشخیص نقاط ویژگی را کنترل می‌کند.  $\lambda$  یک ضریب وزن‌دهی است که اهمیت نسبی جملات را تعیین می‌کند و در این کار برابر با 1 تنظیم شده است.

## نتایج تجربی و شبیه‌سازی‌ها

این بخش، معیارهای ارزیابی برای سنجش روش پیشنهادی را ارائه می‌دهد. ابتدا، معیارهای ارزیابی را تعریف می‌کنیم. سپس، مجموعه داده‌های آموزش و آزمون را معرفی می‌کنیم. در ادامه، جزئیات پیکربندی آموزش را شرح می‌دهیم. در نهایت، با استفاده از داده‌های آزمون، مقایسه جامع و کاملی را در مقایسه با روش‌های دیگر انجام می‌دهیم.

فاصله را دارند. برای دستیابی به این هدف، ما عناصر در  $R$  را بر اساس فاصله آن‌ها از  $d$  رتبه‌بندی می‌کنیم و یک لیست مرتب شده  $L = \{x_1, x_2, \dots, x_N\}$  را تولید می‌کنیم که در آن  $N = |R|$  و  $x_i$  نشان دهنده فاصله بین  $d$  و  $i$  امین توصیف‌گر در  $R$  است. سپس دقت ( $Perc$ ) و فراخوانی ( $Rec$ ) را در هر موقعیت در لیست به صورت زیر محاسبه می‌کنیم:

$$Perc_d(m) = \frac{|L_{i < m} \cap R^+|}{m} \quad (۹)$$

$$Rec_d(m) = \frac{|L_{i < m} \cap R^+|}{|R^+|} \quad (۱۰)$$

که در آن  $L_{i < m}$  مجموعه  $m$  عنصر اول در  $L$  است. با استفاده از این تعاریف، منحنی دقت-فراخوانی (precision-recall) به صورت زیر به دست می‌آید:

$$PR(d) = \{Perc_d(m), Rec_d(m)\}, m = 1, 2, \dots, N \quad (۱۱)$$

میانگین دقت ( $AP$ ) به عنوان مساحت زیر منحنی دقت-فراخوانی محاسبه می‌شود:

$$AP = \sum_{m=1}^N Perc_d(m) \Delta Rec_d(m) = \sum_{m=1}^N Perc_d(m) (Rec_d(m) - Rec_d(m-1)) \quad (۱۲)$$

با این حال، به دلیل ماهیت غیرقابل مشتق‌پذیر عملیات مرتب‌سازی، این تعریف نمی‌تواند مستقیماً بهینه شود. برای رفع این مشکل مرجع [۳۷] با بازتعریف دقت و فراخوانی راه حلی به صورت زیر پیشنهاد کردند:

$$Perc(z) = P(R^+ | Z < z) = \frac{F(Z | R^+) P(R^+)}{F(Z)} \quad (۱۳)$$

$$Rec = P(Z < z | R^+) = F(Z | R^+) \quad (۱۴)$$

که در آن  $Z$  متغیر تصادفی مرتبط با فواصل  $z$  است،  $P(R^+)$  و  $P(R^-) = 1 - P(R^+)$  احتمال‌های پیشین هستند که انحراف فاصله عناصر در  $R$  را نشان می‌دهند، و  $F(z) = P(Z < z)$  تابع توزیع تجمعی (CDF) برای  $Z$  است. با استفاده از این تعاریف،  $AP$  می‌تواند به عنوان  $FastAP$  برای هر توصیف‌گر بازنویسی شود:

$$FastAP = \sum_{z \in Z} \frac{F(Z | R^+) P(R^+)}{F(z)} P(Z | R^+) \quad (۱۵)$$

این فرمولاسیون می‌تواند با استفاده از تقسیم‌بندی هیستوگرام

## معیارهای ارزیابی

برای ارزیابی عملکرد تثبیت تصاویر SAR مجموعه معیارهای ارزیابی زیر در این پژوهش به کار گرفته شده‌اند.

$N_{all}$ : تعداد کل تطابق‌های نقاط کلیدی اولیه بین تصاویر را نشان می‌دهد.

$NOCC$ : تعداد تطابق‌های صحیح میان نقاط ویژه را مشخص می‌کند.

$ROCC$ : نسبت تطابق‌های صحیح به کل تناظرات را اندازه‌گیری می‌کند. هر چه این معیار بزرگ‌تر باشد نشان دهنده تطابق بهتر نقاط کلیدی و حساسیت کمتر به نویز و داده‌های پرت است. این معیار به‌صورت زیر محاسبه می‌شود:

$$ROCC = \frac{NOCC}{N_{all}} \quad (۱۹)$$

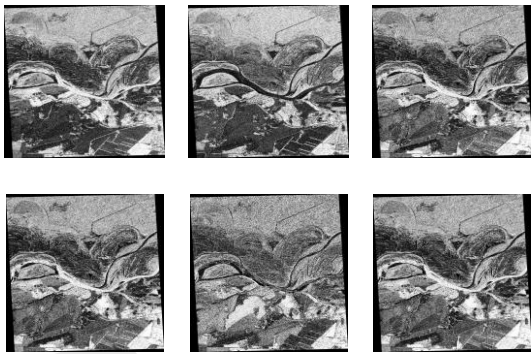
$RMSE$ : ریشه میانگین مربعات خطا، دقت و استحکام تبدیل هندسی بین جفت تصاویر را با استفاده از تبدیل‌های هندسی مستقیم و معکوس ارزیابی می‌کند.  $RMSE$  کمتر نشان دهنده دقت بیشتر در تثبیت است. این معیار به‌صورت زیر محاسبه می‌شود:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \|P_i - f(g(P_i))\|^2} \quad (۲۰)$$

که در آن  $P_i$  یک نقطه ویژه در تصویر اول،  $N$  تعداد کل نقاط ویژه، و  $f$  و  $g$  مدل‌های تبدیل هندسی مستقیم و معکوس هستند.

## داده‌ها

برای آموزش شبکه از مجموعه‌ای از تصاویر از پیش تثبیت شده ماهواره Sentinel-1 استفاده شده است. این تصاویر در Kaggle در دسترس عموم قرار دارند. این مجموعه داده شامل تقریباً ۵۵۰ مکان متنوع جغرافیایی است که شامل زمین‌های بایر، مناطق کوهستانی، مناطق شهری، خطوط ساحلی، بندرها و مناطق کشاورزی می‌شود. نمونه‌هایی از این داده‌ها در شکل ۵ ارائه شده است. برای ارزیابی، یک مجموعه داده آزمون مجزا شامل تصاویر Sentinel-1، Radarsat، ALOS-PALSAR و ERS-2 برای سنجش روش‌های پیشنهادی به کار گرفته شده‌اند. مشخصات جامع داده‌های آزمون در جدول ۱ آمده است. شکل ۶ این داده‌ها را نشان می‌دهد.



شکل ۵- تصاویر نمونه مورد استفاده برای آموزش.

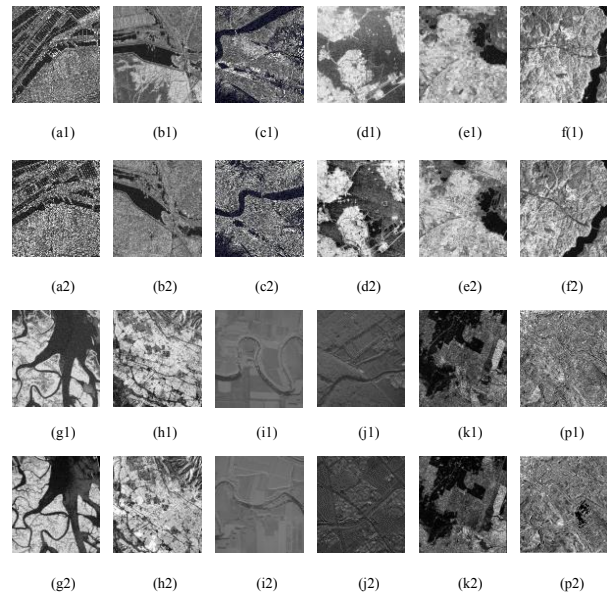
Fig. 5. Sample images used for training.

جدول ۱- مشخصات تصاویر آزمون.

Table 1. Test image specifications.

ID	Sat	Size	Res	Loc	Year
a1	RadarSat-2	400 <sup>2</sup>	8m	China	2009
a2	RadarSat-2	400 <sup>2</sup>	8m	China	2008
b1	RadarSat-2	600×500	8m	China	2009
b2	RadarSat-2	600×500	8m	China	2008
c1	RadarSat-2	400 <sup>2</sup>	8m	China	2009
c2	RadarSat-2	400 <sup>2</sup>	8m	China	2008
d1	Sentinel-1	1024 <sup>2</sup>	10m	Bahrain	2022
d2	Sentinel-1	1024 <sup>2</sup>	10m	Bahrain	2018
e1	Sentinel-1	1024 <sup>2</sup>	10m	Azerbaijan	2022
e2	Sentinel-1	1024 <sup>2</sup>	10m	Azerbaijan	2018
f1	Sentinel-1	1024 <sup>2</sup>	10m	Turkey	2022
f2	Sentinel-1	1024 <sup>2</sup>	10m	Turkey	2018
g1	Sentinel-1	1024 <sup>2</sup>	10m	Iran	2022
g2	Sentinel-1	1024 <sup>2</sup>	10m	Iran	2018
h1	Sentinel-1	1024 <sup>2</sup>	10m	Iran	2022
h2	Sentinel-1	1024 <sup>2</sup>	10m	Iran	2018
i1	Airborne SAR	1597×1602	1.5m	Iran	2020
i2	Airborne SAR	1597×1602	1.5m	Iran	2020
j1	Airborne SAR	1079×1079	1.5m	Iran	2016
j2	Airborne SAR	1079×1079	1.5m	Iran	2016
k1	ALOS-PALSAR	650×350	~	Australia	2018
k2	ALOS-PALSAR	650×350	~	Australia	2019
p1	ERS-2	301 <sup>2</sup>	~	Switzerland	1999
p2	ERS-2	301 <sup>2</sup>	~	Switzerland	1999

تکرارها موفق به تثبیت جفت تصاویر نشدند، با خط خوردگی در نتایج مشخص شده‌اند. روش پیشنهادی، SARNet، در سه حالت تک توصیف‌گری، دو توصیف‌گری و سه توصیف‌گری مورد ارزیابی قرار گرفته است. تعداد زیرتوصیف‌گرها (M) و آستانه نهایی تطابق (N) برای هر مورد مشخص شده است. الگوریتم SIFT ضعیف‌ترین عملکرد را در میان روش‌های مقایسه شده از خود نشان داد. با وجود تعداد تطابق‌های بسیار اندک میان نقاط ویژه، ۸ جفت از ۱۲ تصویر آزمون را تثبیت کرد و برای دو جفت از تصاویر آزمون بالاترین دقت را به دست آورد. الگوریتم SAR-SIFT بهترین عملکرد را برای جفت تصاویر (a1, a2) و (b1, b2) به دست آورد. در حالی که RIFT2 و D2Net تمام موارد را با موفقیت تثبیت کردند و RMSE بزرگتری نسبت به سایر روش‌ها از خود نشان دادند. D2Net در میان رویکردهای یادگیری عمیق دارای تعداد پارامتر بالایی است اما دقت تثبیت به اندازه کافی خوب نیست. شبکه‌های ALIKED و DeDoDe نسبت به روش‌های قبلی در ثبات و دقت عملکرد بهتری داشتند اما نسبت به روش پیشنهادی ما عملکرد ضعیف‌تری از خود نشان دادند. R2D2 در تمام موارد عملکرد مناسبی از خود نشان داد ولی با این وجود، دقت کمتری نسبت به روش‌های پیشنهادی ما داشت. تثبیت جفت‌های (c1, c2)، (f1, f2) و (g1, g2) به دلیل بافت‌های همگن (کناره‌های رودخانه) و ساختارهای تکراری (کوه‌ها) دشوار است، به همین علت اکثر روش‌های به کار رفته دقت بالایی در تثبیت این تصاویر نداشتند. روش پیشنهاد (SARNet) تنها با یک توصیف‌گر، در تثبیت جفت تصاویر ناموفق بوده است. ولی در مابقی موارد دقت خوبی از خود نشان داده است. با استفاده از دو توصیف‌گر و ۴ زیرتوصیف‌گر (N4) و حداقل ۳ تناظر (M3) به صورت میانگین خطای کمتری نسبت به حالت تک توصیف‌گری دارد ولی مشابه حالت قبل در تثبیت جفت تصاویر ناموفق بوده است. در حالت سه توصیف‌گری تصاویر ز نیز با موفقیت تثبیت شده‌اند. و خطا نیز بهبود یافته است. روش پیشنهادی با سه توصیف‌گر و تعداد ۶ زیرتوصیف‌گر و حداقل ۳ تناظر برای الگوریتم ایجاد تناظرات، کمترین خطا را به صورت میانگین پس از روش DAMFE از خود نشان می‌دهد. جدول ۳ نتایج میانگین را در تمام نمونه‌های آزمون نشان می‌دهد. DAMFE با میانگین RMSE برابر با 0.0169 برای تمام تصاویر آزمون بهترین عملکرد را کسب کرد. SARNet در بهترین حالت با RMSE برابر با 0.074 کمی ضعیف‌تر عمل کرد. در بررسی‌های ما، گستره تغییرات خطای تثبیت تصاویر (RMSE) در تکرارهای مختلف برای الگوریتم SIFT بین 0.1 تا 555، برای SAR-SIFT بین 0.01 تا 137 و برای RIFT-2 بین 0.2 تا 35 بوده است. برای روش‌های مبتنی بر یادگیری عمیق از جمله D2Net، ALIKED، DeDoDe و R2D2، غالباً در تمام تکرارها کمتر از 10 بوده است. برای رویکرد DAMFE محدوده تغییرات خطای تثبیت



شکل ۶- تصاویر آزمون مورد استفاده برای ارزیابی.

Fig. 6. Test images used for evaluation.

## تنظیمات آموزش

ما شبکه را با استفاده از بهینه‌ساز Adam، با نرخ یادگیری  $10^{-3}$  و Weight Decay برابر با  $5 \times 10^{-4}$  آموزش دادیم. تصاویر ورودی  $192 \times 192$  پیکسلی با استفاده از تبدیل‌های تصادفی شامل برش، انتقال، مقیاس‌بندی و چرخش که بر روی جفت تصاویر اعمال می‌شدند، تقویت کردیم. آموزش با Batch size ۳ انجام شد. با ۵۵۶ جفت تصویر SAR، هر دوره تقریباً ۵۰ ثانیه بر روی سخت‌افزار زیر طول کشید:

- GPU: NVIDIA Tesla P100 (16 VRAM)
- CPU: Intel(R) Xeon(R) Processor @ 2.00GHz
- RAM: 30GB

این پیکربندی سخت‌افزاری در Kaggle در دسترس بود و ما به مدت ۱۲۰۰ دوره آموزش دادیم.

## مقایسه

روش پیشنهادی را با چندین روش کلاسیک موجود از جمله SIFT، SAR-SIFT و RIFT2 مقایسه کردیم. علاوه بر آن از روش‌های یادگیری عمیق مانند D2Net، ALIKED، DeDoDe و R2D2 نیز در این مقایسه استفاده شده است.

در همه روش‌ها از الگوریتم RANSAC با آستانه خطای ۵ پیکسل و ۵۰۰۰ تکرار برای تخمین مدل هندسی استفاده شده است. به دلیل ماهیت تصادفی RANSAC، هر جفت تصویر ۶۰ بار پردازش شد و بهترین نتایج در جدول ۲ گزارش شدند. مواردی که در هیچ‌کدام از

اگرچه SARNet در دقت از سایر روش‌های مبتنی بر یادگیری پیشی نمی‌گیرد، اما در میان آن‌ها کمترین تعداد پارامتر را دارد. همانطور که در جدول ۴ نشان داده شده است، SARNet تنها از ۱۱۵ هزار پارامتر استفاده می‌کند که به‌طور قابل توجهی کمتر از روش‌های مشابه کمتر است.

حداکثر 15 بوده است. در روش پیشنهادی ما نیز این خطا در تکرارهای مختلف در محدوده‌ای بین 0.01 تا 13 بوده است. لذا می‌توان گفت رویکرد پیشنهادی نسبت به روش‌های کلاسیک پایدارتر و پایداری مشابه رویکردهای مورد مبتنی بر یادگیری مورد بررسی در این پژوهش دارد

**جدول ۲-** نتایج ارزیابی بر روی تصاویر آزمون. N تعداد زیرتوصیف‌گرها را نشان می‌دهد و M حداقل تعداد تناظرات.

**Table 2.** Comparison results over Test images. N denotes to the number of partitions and M to the minimum number of matches.

		a	b	c	d	e	f	g	h	i	j	k	p
SIFT [2]	$N_{all}$	559	656	382	615	730	767	487	750	412	575	689	534
	ROCC	0.055	0.027	0.094	0.011	0.149	0.027	0.023	0.073	0.344	0.01	0.05	0.009
	RMSE	1.369	1.064	0.12	455	0.001	0.951	0.694	0.002	0.007	316	0.297	298
SAR-SIFT [13]	$N_{all}$	175	229	151	367	606	906	518	1425	37	127	790	545
	ROCC	0.509	0.686	0.503	0.114	0.469	0.243	0.214	0.298	0.783	0.110	0.631	0.447
	RMSE	0.021	0.009	0.033	0.761	0.157	0.216	0.523	0.241	0.006	0.815	0.114	0.092
RIFT2 [11]	$N_{all}$	763	735	441	1279	2081	2071	943	2019	1153	4092	1264	570
	ROCC	0.505	0.103	0.113	0.030	0.066	0.032	0.038	0.055	0.095	0.944	0.176	0.528
	RMSE	0.267	0.344	0.585	2.151	1.835	3.154	2.602	2.025	0.755	0.009	0.372	0.236
D2Net[18]	$N_{all}$	294	382	339	1990	3230	2510	1669	3506	3485	1283	1079	1315
	ROCC	0.531	0.466	0.572	0.456	0.634	0.269	0.318	0.62	0.475	0.116	0.846	0.505
	RMSE	0.120	0.228	0.133	0.048	0.114	0.166	0.168	0.66	0.125	3.88	0.053	0.092
ALIKED [20]	$N_{all}$	247	340	125	1895	2401	1920	1749	2238	1653	1576	488	1860
	ROCC	0.526	0.347	0.368	0.306	0.616	0.307	0.178	0.548	0.310	0.041	0.485	0.375
	RMSE	0.105	0.130	0.021	0.214	0.068	0.104	0.638	0.089	0.077	2.772	0.120	0.102
DeDoDe [21]	$N_{all}$	126	152	14	656	1233	224	429	1807	1211	1752	890	1034
	ROCC	0.786	0.783	0.857	0.787	0.917	0.647	0.531	0.855	0.76	0.83	0.921	0.755
	RMSE	0.037	0.123	0.018	0.114	0.04	0.132	0.495	0.114	0.147	0.921	0.027	0.157
R2D2[19]	$N_{all}$	106	206	103	896	1133	704	713	1167	1663	531	332	738
	ROCC	0.160	0.101	0.116	0.434	0.556	0.049	0.072	0.544	0.523	0.075	0.635	0.285
	RMSE	0.030	0.732	0.131	0.003	0.009	0.588	0.947	0.004	0.017	0.945	0.019	0.006
DAMFE[38]	$N_{all}$	52	66	55	64	73	72	62	80	74	34	65	75
	ROCC	0.423	0.393	0.36	0.546	0.767	0.61	0.274	0.575	0.79	0.73	0.476	0.37
	RMSE	0.029	0.023	0.05	0.008	0.004	0.013	0.008	0.006	0.008	0.023	0.019	0.012
SARNet (Single Desc)	$N_{all}$	418	834	429	2787	2827	2491	2286	2670	5515	2003	1150	2455
	ROCC	0.127	0.128	0.182	0.211	0.357	0.056	0.038	0.326	0.176	0.003	0.542	0.124
	RMSE	0.042	0.328	0.206	0.052	0.034	0.511	3.277	0.056	0.014	357	0.107	0.307
SARNet (Dual Desc) (N4, M3)	$N_{all}$	29	35	30	155	252	59	72	184	318	28	382	131
	ROCC	0.517	0.543	0.7	0.8	0.881	0.136	0.139	0.864	0.827	0.143	0.94	0.588
	RMSE	0.066	0.09	0.065	0.002	0.028	0.025	0.007	0.003	0.005	591	0.057	0.08

SARNet (N3, M3)	$N_{all}$	256	509	286	1901	2029	1625	1594	2008	3672	1239	822	1576
	ROCC	0.172	0.124	0.21	0.26	0.466	0.079	0.061	0.407	0.24	0.097	0.554	0.132
	RMSE	0.212	0.470	0.023	0.055	0.039	0.496	1.24	0.082	0.054	0.624	0.140	0.126
SARNet (N6, M3)	$N_{all}$	83	127	90	422	618	269	310	517	808	703	562	328
	ROCC	0.386	0.354	0.467	0.59	0.709	0.078	0.087	0.70	0.608	0.074	0.831	0.405
	RMSE	0.060	0.070	0.027	0.003	0.025	0.014	0.428	0.020	0.056	0.053	0.097	0.047

جدول ۳- میانگین نتایج بر روی تمام تصاویر آزمون.

Table 3. Average results over all test samples.

	SIFT [2]	SAR-SIFT [13]	RIFT2 [11]	D2Net[18]	ALIKED [20]	DeDoDe [21]
$N_{all}$	596.333	459.08	1450.917	1756.833	1374.333	787.5
ROCC	0.072	0.417	0.223	0.484	0.367	0.796
RMSE	89.458	0.249	1.1945	0.482	0.370	0.184
	R2D2[19]	DAMFE[39]	SARNet (Single Desc)	SARNet (Dual Desc) (N4,M3)	SARNet (N3, M3)	SARNet (N6, M3)
$N_{all}$	691	64.333	2155.417	139.583	1459.75	403.083
ROCC	0.295	0.526	0.189	0.589	0.272	0.499
RMSE	0.285	0.0169	30.161	49.285	0.296	0.075

جدول ۴- مقایسه‌ی تعداد پارامترهای شبکه‌های عصبی عمیق.

Table 4. Comparison of parameters count in different networks.

Models	Million Parameters
D2-Net [18]	7.635
LF-Net [39]	2.642
SuperPoint [40]	1.301
R2D2[19]	0.484
ASLFeat [41]	0.823
DISK [42]	1.092
ALIKE-N [43]	0.318
ALIKE-L [43]	0.653
DeDoDe (Detector) [21]	14.559
DeDoDe (Descriptor) [21]	13.518
ALIKED-T(16) [20]	0.192
ALIKED-N(32) [20]	0.677
ALIKED-N(32) [20]	0.980
DAMFE[38]	0.484
SARNet	0.115

## نتیجه گیری

در این مقاله، به چالش تثبیت تصاویر SAR پرداختیم که به دلیل ویژگی‌های منحصر به فرد خود از جمله نویز لکه‌ای و بافت‌های همگن پیچیدگی‌های خاصی دارد. در این مقاله روشی برای تثبیت این تصاویر معرفی کردیم. شبکه SARNet، یک شبکه عصبی عمیق سبک است که سه توصیف‌گر متمایز را برای هر نقطه کلیدی تولید می‌کند. این شبکه بر روی تصاویر SAR آموزش داده و ارزیابی شد. نتایج ارزیابی‌ها قابل مقایسه با روش‌های جدید است. روش پیشنهادی به دلیل تعداد پارامترهای کم (۱۱۵ هزار) خود متمایز است، و سبک‌ترین شبکه در میان روش‌های یادگیری عمیق مقایسه شده در این پژوهش است. این ویژگی به‌ویژه برای استقرار در محیط‌هایی با منابع محدود مزیت بزرگی است. این شبکه به‌طور دقیق بر روی یک مجموعه داده متنوع شامل ماهواره‌های متعدد و مناطق مختلف جغرافیایی ارزیابی شد. نتایج شبیه‌سازی‌ها، اثربخشی رویکرد ما را تأیید می‌کنند.

## تعارض منافع

هیچگونه تعارض منافع توسط نویسندگان بیان نشده است.

## منابع

- pp. 346-359, 2008, <https://doi.org/10.1016/j.cviu.2007.09.014>.
- [7] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "KAZE features," in *Computer Vision–ECCV 2012, 12<sup>th</sup> European Conference on Computer Vision Florence, Italy, 2012*, A. Fitzgibbon, S. Lazebnik, P. Perona, and Y. Sato, Eds. Berlin, Heidelberg: Springer, 2012, pp. 214-227, [https://doi.org/10.1007/978-3-642-33783-3\\_16](https://doi.org/10.1007/978-3-642-33783-3_16).
- [8] P. F. Alcantarilla, J. Nuevo, and A. Bartoli, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," in *British Machine Vision Conference (BMVC)*, Bristol, UK, 2013, <https://doi.org/10.5244/C.27.13>.
- [9] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *International Conference on Computer Vision, Barcelona, Spain, 2011*, pp. 2564-2571, <https://doi.org/10.1109/ICCV.2011.6126544>.
- [10] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary," in *Computer Vision–ECCV 2010, 11<sup>th</sup> European Conference on Computer Vision, Heraklion, Crete, Greece, 2010, Part IV*, K. Daniilidis, P. Maragos, and N. Paragios, Eds. Berlin, Heidelberg: Springer, 2010, pp. 778-792, [https://doi.org/10.1007/978-3-642-15561-1\\_56](https://doi.org/10.1007/978-3-642-15561-1_56).
- [11] J. Li, P. Shi, Q. Hu, and Y. Zhang, "RIFT2: Speeding-up RIFT with a new rotation-invariance technique," *Journal of Latex Class Files*, vol. 14, no. 8, pp. 1-6, 2021, <https://doi.org/10.48550/arXiv.2303.00319>.
- [12] Z. Hou, Y. Liu, and L. Zhang, "POS-GIFT: A geometric and intensity-invariant feature transformation for multimodal images," *Information Fusion*, vol. 102, 2024, Art. no. 102027, <https://doi.org/10.1016/j.inffus.2023.102027>.
- [13] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, and F. Tupin, "SAR-SIFT: a SIFT-like algorithm for SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 1, pp. 453-466, 2015, <https://doi.org/10.1109/TGRS.2014.2323552>.
- [14] W. Ma, Z. Wen, Y. Wu, L. Jiao, M. Gong, and Y. Zheng, "Remote sensing image registration with modified SIFT and enhanced feature matching," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 1, pp. 3-7, 2016, <https://doi.org/10.1109/LGRS.2016.2600858>.
- [15] M. Pourfard, T. Hosseinian, R. Saeidi, A. A. Motamedi, M. J. Abdollahifard, and R. Mansoori, "KAZE-SAR: SAR image registration using KAZE detector and modified SURF Descriptor for tackling speckle noise," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, 2021, Art. no. 5207612, <https://doi.org/10.1109/TGRS.2021.3084411>.
- [16] Z. Xie, W. Zhang, L. Wang, J. Zhou, and Z. Li, [1] A. Tsokas, M. Rysz, P. M. Pardalos, and K. Dipple, "SAR data applications in earth observation: An overview," *Expert Systems with Applications*, vol. 205, 2022, Art. no. 117342, <https://doi.org/10.1016/j.eswa.2022.117342>.
- [2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004, <https://doi.org/10.1023/B:VISI.0000029664.99615.94>.
- [3] J. M. Morel and G. Yu, "ASIFT: A new framework for fully affine invariant image comparison," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 438-469, 2009, <https://doi.org/10.1137/080732730>.
- [4] Z. S. Ni, "B-SIFT: A Binary SIFT based local image feature descriptor," in *Fourth International Conference on Digital Home*, Washington, DC, United States, 2012, pp. 117-121, <https://doi.org/10.1109/ICDH.2012.69>.
- [5] R. Arandjelović and A. Zisserman, "Three things everyone should know to improve object retrieval," in *Conference On Computer Vision and Pattern Recognition*, Providence, RI, USA, 2012, pp. 2911-2918, <https://doi.org/10.1109/CVPR.2012.6248018>.
- [6] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-Up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3,

- and Remote Sensing, vol. 63, 2025, Art. no. 5214918, <https://doi.org/10.1109/TGRS.2025.3575193>.
- [27] H. Fang, Q. Meng, G. Liao, Y. Liu, C. Zeng, and Z. Chen, "A simple and effective deep registration algorithm for SAR images," in *7<sup>th</sup> International Conference on Information Communication and Signal Processing (ICICSP)*, Zhoushan, China, 2024, pp. 191-195, <https://doi.org/10.1109/ICICSP62589.2024.10809248>.
- [28] X. Hu, Y. Wu, Z. Li, Z. Yang, and M. Li, "Multi-feature alignment and matching network for SAR and optical image registration," *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 18, pp. 352-367, 2024, <https://doi.org/10.1109/JSTARS.2024.3492278>.
- [29] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is "nearest neighbor" meaningful?," in *Database Theory—ICDT'99, 7<sup>th</sup> International Conference Jerusalem, Israel, 1999*, C. Beeri and P. Buneman, Eds. Berlin, Heidelberg: Springer, 1999, pp. 217-235, [https://doi.org/10.1007/3-540-49257-7\\_15](https://doi.org/10.1007/3-540-49257-7_15).
- [30] H. Liu and S. Zhang, "Noisy data elimination using mutual k-nearest neighbor for classification mining," *Journal of Systems and Software*, vol. 85, no. 5, pp. 1067-1074, vol. 85, no. 5, pp. 1067-1074, 2012, <https://doi.org/10.1016/j.jss.2011.12.019>.
- [31] M. Muja and D. G. Lowe, "Scalable nearest neighbor algorithms for high dimensional data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2227-2240, 2014, <https://doi.org/10.1109/TPAMI.2014.2321376>.
- [32] J. Norouzi, M. S. Helfroush, A. Liaghat, and H. Danyali, "Beyond outlier removal: Integrated ensemble matching for accurate image keypoint correspondence," *Knowledge-Based Systems*, vol. 316, 2025, Art. no. 113343, <https://doi.org/10.1016/j.knosys.2025.113343>.
- [33] O. Chum and J. Matas, "Optimal randomized RANSAC," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 8, pp. 1472-1482, 2008, <https://doi.org/10.1109/TPAMI.2007.70787>.
- [34] R. Raguram, O. Chum, M. Pollefeys, J. Matas, and J. M. Frahm, "USAC: A universal framework for random sample consensus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 2022-2038, 2012, <https://doi.org/10.1109/TPAMI.2012.257>.
- [35] P. E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superglue: Learning feature matching with graph neural networks," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2020, pp. 4938-4949, <https://doi.org/10.1109/CVPR42600.2020.00499>.
- [36] P. Lindenberger, P. E. Sarlin, and M. Pollefeys, "Optical and SAR image registration based on the phase congruency framework," vol. 13, no. 10, 2023, Art. no. 5887, <https://doi.org/10.3390/app13105887>.
- [17] Y. Verdie, K. M. Yi, P. Fua, and V. Lepetit, "TILDE: A temporally invariant learned detector," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 5279-5288, <https://doi.org/10.1109/CVPR.2015.7299165>.
- [18] M. Dusmanu et al., "D2-Net: A trainable CNN for joint description and detection of local features," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 8084-8093, <https://doi.org/10.1109/CVPR.2019.00828>.
- [19] J. Revaud, C. De Souza, M. Humenberger, and P. Weinzaepfel, "R2d2: Reliable and repeatable detector and descriptor," in *33<sup>rd</sup> International Conference on Neural Information Processing Systems*, Vancouver, Canada, 2019, Art. no. 1113, pp. 12414-12424.
- [20] X. Zhao, X. Wu, W. Chen, P. C. Chen, Q. Xu, and Z. Li, "Aliked: A lighter keypoint and descriptor extraction network via deformable transformation," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, 2023, Art. no. 5014016, <https://doi.org/10.1109/TIM.2023.3271000>.
- [21] J. Edstedt, G. Bökman, M. Wadenbäck, and M. Felsberg, "DeDoDe: Detect, don't describe- describe, don't detect for local feature matching," in *International Conference on 3D Vision (3DV)*, Davos, Switzerland, 2024, <https://doi.org/10.1109/3DV62453.2024.00035>.
- [22] Z. Li, H. Zhang, and Y. Huang, "A rotation-invariant optical and SAR image registration algorithm based on deep and gaussian features," *Remote Sensing*, vol. 13, no. 13, 2021, Art. no. 2628, <https://doi.org/10.3390/rs13132628>.
- [23] L. H. Hughes, D. Marcos, S. Lobry, D. Tuia, and M. Schmitt, "A deep learning framework for matching of SAR and optical imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 169, pp. 166-179, 2020, <https://doi.org/10.1016/j.isprsjprs.2020.09.012>.
- [24] D. Quan, H. Wei, R. Lei, B. Duan, and Y. Li, "Self-distillation feature learning network for optical and SAR image registration," *Transactions on Geoscience and Remote Sensing*, vol. 60, 2022, Art. no. 4706718, <https://doi.org/10.1109/TGRS.2022.3173476>.
- [25] S. Mao et al., "Cross-rejective open-set SAR image registration," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 2025, pp. 23027-23036, <https://doi.org/10.1109/CVPR52734.2025.02144>.
- [26] A. Liaghat, M. S. Helfroush, J. Norouzi, H. Danyali, and J. M. Guerrero, "OS-CycleGAN: modified CycleGAN-based descriptors for optical and SAR image matching," *IEEE Transactions on Geoscience*

- [40]D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superpoint: Self-supervised interest point detection and description," in *Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, UT, USA, 2018, pp. 224-236 ., <https://doi.org/10.1109/CVPRW.2018.00060>.
- [41]Z. Luo *et al.*, "Aslfeat: Learning local features of accurate shape and localization," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2020, pp. 6589-659, <https://doi.org/10.1109/CVPR42600.2020.00662>.
- [42]M. Tyszkiewicz, P. Fua, and E. Trulls, "DISK: Learning local features with policy gradient," in *34nd Conference on Neural Information Processing Systems (NeurIPS 2020)*, Vancouver, Canada. 2018.
- [43]X. Zhao, X. Wu, J. Miao, W. Chen, P. C. Y. Chen, and Z. Li, "ALIKE: Accurate and lightweight keypoint detection and descriptor extraction," *IEEE Transactions on Multimedia*, vol. 25, pp. 311-3112, 2022, <https://doi.org/10.1109/TMM.2022.3155927>.
- "Lightglue: Local feature matching at light speed," in *International Conference on Computer Vision (ICCV)*, Paris, France, 2023, pp. 17627-17638, <https://doi.org/10.1109/ICCV51070.2023.01616>.
- [37]F. Cakir, K. He, X. Xia, B. Kulis, and S. Sclaroff, "Deep metric learning to rank," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 1861-1870, <https://doi.org/10.1109/CVPR.2019.00196>.
- [38]J. Norouzi, M. S. Helfroush, A. Liaghat, and H. Danyali, "A deep-based approach for multi-descriptor feature extraction: Applications on SAR image registration," *Expert Systems with Applications*, vol. 254, 2024, Art. no. 124291, <https://doi.org/10.1016/j.eswa.2024.124291>.
- [39]Y. Ono, E. Trulls, P. Fua, and K. M. Yi, "LF-Net: Learning local features from images," in *32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*, Montréal, Canada. 2018.