Marine Pollution Bulletin

Increasing the Sentinel-2 potential for marine plastic litter monitoring through image fusion techniques

Maria Kremezi[a]*, Viktoria Kristollari[a], Vassilia Karathanassi[a], Konstantinos Topouzelis[b], Pol Kolokoussis[a], Nicolò Taggio[c], Antonello Aiello[c], Giulio Ceriola[c], Enrico Barbone[d] and Paolo Corradi[e]

[a]*Laboratory of Remote Sensing, National Technical University of Athens, School of Rural, Surveying, and Geoinformatics Engineering, Zografou 15780, Greece*
[b]*Department of Marine Sciences, University of the Aegean, Mytilene 81100, Greece*
[c]*Planetek Italia, Bari 70132, Italy*
[d]*ARPA Puglia, Environmental Protection Agency of Puglia Region, Bari 70126, Italy*
[e]*European Space Research and Technology Centre (ESTEC), European Space Agency, Noordwijk 2200 AG, Netherlands*

Abstract
Sentinel-2 (S2) images have been used in several projects to detect large accumulations of marine litter and plastic targets. Their limited spatial resolution though hinders the detection of relatively small floating accumulations of marine debris. Thus, this study aims at overcoming this limit through the exploration of fusion with very high-resolution WorldView-2/3 (WV-2/3) images. Various state-of-the-art approaches (component substitution, spectral unmixing, deep learning) were applied on data collected in synchronized acquisitions of plastic targets of various sizes and materials in seawater. The fused images were evaluated for spectral and spatial distortions, as well as their ability to spectrally discriminate plastics from water. Several WV-2/3 band combinations were investigated and five litter indexes were applied. Results showed that: a) the VNIR combination is the optimal one, b) the smallest observable plastic target is 0.6×0.6 m^2 and c) SWIR bands are important for marine litter detection.

Keywords: satellite data, image fusion, marine pollution, plastic litter detection, controlled experiments, spectral analysis.

1. Introduction

Marine litter is defined as any persistent, manufactured or processed solid material discarded, disposed of or abandoned in the marine and coastal environment (UNEP, 2005). Marine litter may be found originating both on land (i.e. river discharges, flood water events, industrial, and recreational littering, etc.) and at sea (fishing, aquaculture, offshore mining and extraction operations, etc.). Land-based activities account for roughly 80% of marine litter (Jambeck *et al*. 2015). Plastics are the most prevalent debris found due to an increase in demand and production of plastic items over the last 70 years, as well as their slow decomposition (Chamas *et al.* 2020; Smail *et al.* 2019). High concentrations of marine litter endanger marine wildlife through entanglement, colonization of surface areas, or ingestion. The latter negatively impacts human health as well as the marine wildlife which is part of our food chain (Rios *et al.* 2007; Casabianca *et al*. 2019; GESAMP 2019). In addition to harming marine life, marine litter has a broad range of negative environmental, socio-economic and maritime travel safety impact (Rios *et al.* 2007; Casabianca *et al.* 2019; GESAMP 2019). Although plastic litter has been reported since the 1960s, it has become a global environmental concern only in more recent years. Scientists estimate that by 2050 the plastic litter mass will overweigh the mass of the fish population (Salgado-Hernanz et al., 2021).

Satellite remote sensing has been identified as a useful tool for marine debris monitoring as it provides global and continuous temporal coverage (Maximenko *et al.* 2019; UNEP, 2016). However, it imposes some significant challenges in terms of atmospheric and sea-surface effects, radiometric, spatial and temporal resolutions, and availability of ground-truth data (Topouzelis *et al.* 2021). Acuña-Ruz *et al.* (2018) created a spectral library from laboratory spectroscopic measurements and used WV-3 images from 0.3 to 1.2 m spatial resolution to train machine learning methods to identify Anthropogenic Marine Debris (AMD) mixtures over beaches on Chiloé island. Their effort to distinguish the various plastic materials showed that while the laboratory analysis revealed a) distinctive absorption patterns in the SWIR (Shortwave Infrared) wavelength region, b) peak reflectance in the NIR (Near Infrared) spectrum, and c) variability in the VIS (visible) spectrum, the WV-3 spectral resolution could only distinguish Expanded Polystyrene (EPS) from the other marine debris. Garaba and Dierssen (2018) used a lab spectral library of various plastic materials as reference for the detection of a large-sized ghost net in airborne SASI (Shortwave infrared Airborne Spectrographic Imager) SWIR imagery with pixel size 0.5×1.2 m^2. They observed that both wet and dry plastic spectra have absorption features around 1215 and 1732 nm, and the reflectance of wet plastic is lower compared to the one of dry plastic, due to water absorption. This highlights the effect of ocean water on the spectral properties of polymers

* Corresponding author.
E-mail address: mkremezi@central.ntua.gr (M. Kremezi)

and the need to investigate the minimum plastic coverage of the ground pixel area that can still allow detection. Lower reflectance for wet plastic spectra was also shown by Goddijn-Murphy and Dufaur (2018) who verified a theoretical model of plastic reflectance. Their model however produced smaller reduction compared to Garaba and Dierssen (2018), a fact that could be explained by the differences in the experimental design and the properties of the selected plastic objects. Topouzelis *et al.* (2019) were the first to conduct an experiment with artificial floating targets consisting of 10×10 m^2 PET (Polyethylene terephthalate)-1.5 L water bottles, LDPE (Low-density Polyethylene) plastic bags, and nylon fishing ghost nets. Sentinel-2 (S2) and Sentinel-1 images were investigated. In the optical data, all 10×10 m^2 plastic targets were distinguishable from the water due to their higher reflectance. In SAR data though, only the plastic bottles could be detected. In a follow-up study (Topouzelis *et al.* 2020), S2 images of artificial plastic targets deployed at sea were examined in combination with UAV (Unmanned Aerial Vehicle) optical multispectral data. Pixel coverage of plastic and linear spectral mixture were used to modify the spectra provided by the USGS spectral library. Then, matched filtering was used to classify the pixels containing concentrations of plastic. This classification technique proved very promising as marine plastic litter targets were successfully identified if they covered at least 25% of the ground sampling distance (GSD).

Biermann *et al.* (2020) made use of S2 images of various locations where marine plastic litter has been reported on scientific publications, mass, or social media. NDVI and the Floating Debris Index (FDI), which was developed in the study of Biermann *et al.* (2020), were used to detect floating litter patches and discriminate plastic from other materials. Both indexes were employed to train a naïve Bayes classifier to compute the probability of a detected pixel to belong to each of the following classes: seaweed, spume, timber, macroplastics, and seawater. The algorithm classified plastic with an overall accuracy of 86%. Martinez-Vicente *et al.* (2020) developed the Normalised Difference Hydrocarbon Index (NDHI) for the detection of plastic on the shoreline using airborne hyperspectral (HS) data. This index is based on the Hydrocarbon Index (HI) (Kühn *et al.* 2004). The results were promising, indicating that subpixel detection is possible, while further investigation is needed to determine the minimum percentage of coverage that can be detected. *Basu et al.* (2021) evaluated the potential of supervised (SVR (Support Vector Regression) and SFCM (Semi-supervised Fuzzy c-means)) and unsupervised (k-means and FCM (Fuzzy c-means)) classification algorithms to detect floating marine litter. The authors used S2 images containing floating marine litter targets with sizes 10×10 m^2 and 3×10 m^2, and various combinations of bands and indexes as attribute sets for the classification algorithms. The supervised classification yielded higher accuracy, while the unsupervised algorithms provided many misclassifications. From the above, it was concluded that when artificial large-sized plastic targets are used, S2 spectral resolution (13 bands in the 440-2200 nm part of the spectrum) can detect marine plastic litter.

In order to increase PRISMA hyperspectral satellite image ability to detect small sized plastic targets, Kremezi *et al.* (2021) evaluated several pan-sharpening methods on PRISMA HS images with 30 m native spatial resolution, based on the ability of the methods to spectrally differentiate plastics from water with the minimum spatial distortions. Through controlled experiments with plastic targets of various sizes, it was observed that in order for a small plastic target to be distinguishable it should at least occupy 8% of a pixel of the original HS image. Moreover, they proposed a combination of three indexes to detect plastic objects.

Spatial resolution, multispectral characteristics (namely number, position, and width of the acquisition bands) and signal-to-noise ratio (SNR) are crucial factors in the design of a sensor dedicated to the detection and discrimination of accumulations of marine litter from space. Current orbiting sensors were not designed for such an application. As above summarised, a number of studies exploited the spatial resolution and multispectral characteristics of S2 to perform detection and discrimination of marine litter accumulations (and targets), however it is now clear that more spatial resolution and more spectral bands, arranged in a dedicated configuration, would significantly improve the detection of marine litter accumulations from orbit. The possibility to fuse images of different orbiting sensors in view to improve at least one of these aspects, would be consequently beneficial.

Thus, in this study, we evaluate various state-of-the-art image fusion algorithms that make use of component substitution, spectral unmixing or deep learning (DL), on S2 and WV-2/3 (2, 4 m spatial resolution) datasets. The DL literature networks were adjusted to the fusion problem since they originated from either the pan-sharpening or the single image super-resolution (SISR) domain. In addition, three DL networks were created for the purpose of the study. Finally, we experiment with various WV-2/3 band combinations to find the optimal one and examine various indexes for their capability to detect floating plastic objects on the fused images.

2. Materials and methods

2.1. Data description

Two controlled experiments with artificial plastic targets were conducted. The first processes S2 and WV-3 images, while the second S2 and WV-2 images. Experiments took place in the Tsamakia beach and the Geras Gulf, respectively (Figure 1). Both locations are found in the coastal region of Lesvos Island, Greece, they are protected from any human activities and they provide conditions of both shallow and deep water.

2.1.1. Field data

The first experiment utilised three 10×10 m^2 plastic targets constructed for the needs of the "Plastic Litter Project 2018" conducted by the Marine Remote Sensing Group of the University of Aegean on June 7th, 2018. Each target contained a different plastic material: PET-1.5 l bottles, LDPE bags, and yellowish nylon fishing nets (Figure 1). All the targets are visible in the WV-3 image while only the “plastic bottles” and “fishing nets” are visible in the S2 image (Figure 2). The

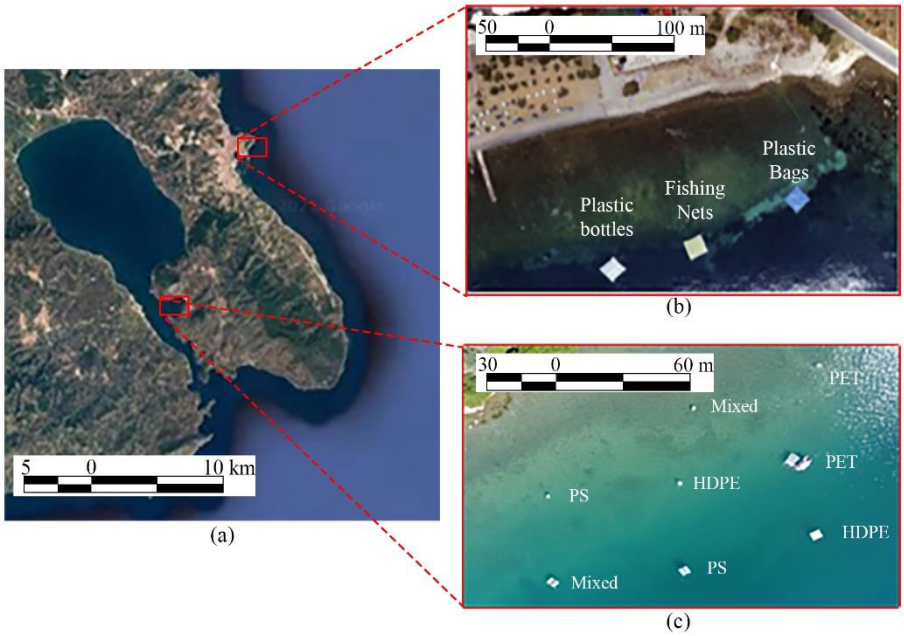

Fig. 1. (a) Google earth image with the test areas highlighted, (b) UAV photograph of the plastic targets of the first experiment (Source: Topouzelis *et al.* 2019) and (c) UAV photograph showing the 8 plastic targets of the second experiment.

second experiment was conducted on June 21st, 2021 utilising 8 floating plastic targets in two dimensions: 4 plastic targets with dimensions 2.4×2.4 m^2 and 4 plastic targets with dimensions 0.6×0.6 m^2 (Figure 1) (Kremezi *et al.* 2021). These targets were constructed using 4 plastic materials: 1) High-density PE (HDPE) (tarps in white, yellow, and green colour), 2) PET (transparent water bottles and green oil bottles), 3) Polystyrene (PS) (building insulation material in cyan colour), and 4) all the above materials in equal quantity. Furthermore, two circular targets were placed in the gulf with a radius of 28 meters. The first contained a white HDPE mesh (PLP, 2021), while the other was constructed entirely from wood. All but the 0.6 m PET targets are visible in the WV-2 image, while only the two circular 28 m targets are visible in the S2 image (Figure 2). All of the constructed targets were anchored in place for both the S2 and WV-2/3 satellite acquisitions.

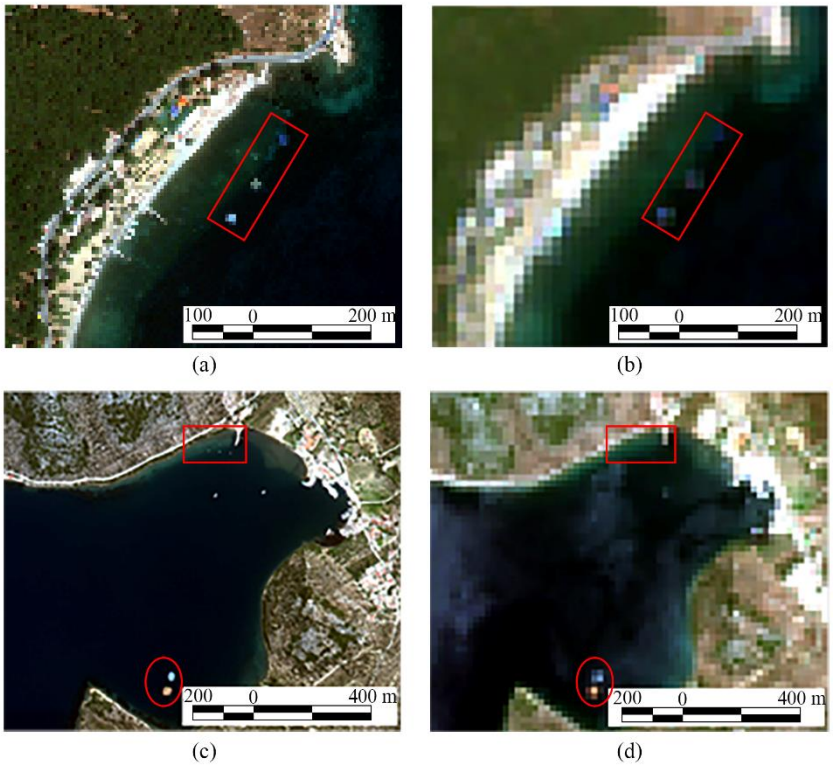

Fig. 2. (a) Natural colours RGB composite of WV-3 image with 4 m spatial resolution and (b) S2 image with 10 m spatial resolution acquired on 07/06/2018. (c) Natural colours RGB composite of WV-2 image with 2 m spatial resolution, and (d) S2 image with 20m spatial resolution acquired on 21/06/2021. Red rectangles indicate the area where square plastic targets are located. Red ellipses indicate the area with the circular plastic and wooden targets.

2.1.2. Satellite data

The data fusion methods were applied on S2 and WV-2/3 satellite images. S2 carries an optical instrument payload (MSI) that samples 13 spectral bands (442-2200 nm): 4 bands at 10 m, 6 bands at 20 m, and three bands at 60 m spatial resolution, while WV-3 provides panchromatic imagery (450-800 nm) with 0.31 m resolution, 8-band Visible-Near Infrared (VNIR) imagery (400-1040 nm) with 1.24 m resolution, and 8-band shortwave infrared (SWIR) imagery (1195-2365 nm) with 3.7 m resolution. WV-2 imagery includes a panchromatic image with 0.46 m resolution and 8 bands at the VNIR region in the same wavelengths as WV-3 with 1.8 m resolution. All WV spatial resolutions that have been mentioned above refer to nadir captured imagery and worsen for higher looking angles of the sensor. In this study, we consider S2 images as the images with high spectral resolution and the WV-2/3 images as the images with high spatial resolution.

2.1.3. Pre-processing

Two S2 L1C images were collected on June 7th, 2018 and June 21st, 2021, respectively. For the experiment on June 7th, 2018, a WV-3 image was acquired almost synchronously with the S2 data, while for the experiment on June 21st, 2021, the WV-2 satellite captured the area at a time close to the S2 sensing time (Figure 2). On both dates of image acquisitions over the test areas, the targets were located offshore and there were clear sky and calm sea conditions. The procured WV-2/3 images were not atmospherically corrected by the commercial provider, but the datasets were converted to Top-Of-Atmosphere (TOA) reflectance to be comparable with the S2 datasets. This task was carried out by using the instructions from the product provider (Kuester, 2016; Updike and Comp, 2010).

Atmospheric corrections were not applied because of a) the clear sky conditions during the experiments, b) concerns expressed in the literature about the possible reduction effects on the spectral signatures (Topouzelis *et al.* 2019; Papageorgiou 2019) which could weaken the plastic signal to undetectable levels, c) the differences in the output between various algorithms (Biermann *et al.* 2020), and d) the low sensitivity of band subtraction indexes to environmental conditions (Hu 2009; Hu *et al.* 2012). In addition, mitigation strategies were not applied for the effects of the different observational geometries (BRDF) since issues that could affect the performance of the fusion process were not observed (e.g. light anisotropies, sunglint) (Kremezi and Karathanassi, 2019).

Fusion approaches utilised all 13 bands of the S2 images, resampled at 20 m spatial resolution. This sampling size was chosen because almost all the bands in NIR and SWIR have 20 m resolution. WV-3 images were resampled at 4 m spatial resolution, while WV-2 images at 2 m. For the DL methods 8 VNIR + 2 SWIR WV bands were used, and for the non-DL, 7 VNIR + 2 SWIR (coinciding WV-S2 bands). The georeference of all datasets was checked to ensure alignment between the S2 and WV-2/3 images. Further co-registration steps were not considered necessary.

2.2. Fusion methods

2.2.1 Related work

Image fusion is the optimal solution to the technological limitations of the spatial and spectral resolutions of a satellite sensor. Image fusion techniques evolved from multispectral (MS) pan-sharpening [MS + PAN (panchromatic) fusion] to MS + MS and MS + HS image fusion (Loncan *et al.* 2015). The first approach was developed by adapting various pan-sharpening/hyper-sharpening techniques to the HS+MS/MS+MS image fusion problem (Component Substitution, Multiresolution Analysis, Sparse representation) (Yokoya *et al.* 2017).

Another more popular category of such methods are algorithms based on spectral unmixing. They exploit all the bands of the common spectral region of the initial images (Bieniarz *et al.* 2011; Lanaras *et al.* 2017; Wei *et al.* 2016; Simoes *et al.* 2015), and they are ideally used on HS datasets since the high spectral resolution is necessary for decomposing the mixed pixels, but they have also been proven useful for MS datasets. The spectral unmixing approach has been used in the HS+MS/MS+MS fusion problem for the last two decades. More recently, Berné *et al.* (2010) implemented the fusion process by unmixing the low-resolution HS image using the non-negative matrix factorization (NMF) algorithm and then obtained the high-resolution abundance maps from the MS image by least squares regression. Yokoya *et al.* (2011) presented the coupled NMF (CNMF) algorithm where the endmember and abundance matrices are estimated alternatively by spectral unmixing based on NMF, taking into account the sensors' spectral response functions (SRFs) and point spread functions (PSFs). Akhtar *et al.* (2015) learned distributions for the scene spectra and their proportions in the image using the Bayesian dictionary learning and used this information to sparse code the high spatial resolution image. Lanaras *et al.* (2015) suggested a projected gradient method into the alternate updates of the endmember and abundances matrices. Wei *et al.* (2015) formulated the fused image reconstruction problem as an ill-posed linear inverse so they included a regularization term to convert it into a well-posed inverse problem (Bayesian approach). In addition they employed an alternating optimization of fusion and unmixing with Sylvester equation solvers which significantly decreased the computational complexity (Wei *et al.* 2015b). Finally, Simoes et *al.* (2015) included an abundance regularization to achieve spatial smoothness of the abundances and ensure that the fused image retains both the spatial and spectral correlation.

In recent years, several super-resolution DL methods have been proposed to super-resolve S2 images. In their vast majority, current research has proposed SISR methods based on GAN (Generative Adversarial Network) (Goodfellow *et al.* 2014) and ResNet (He *et al.* 2016) architectures. Although our study uses two sources of data, SISR DL networks can be easily adjusted to fit the fusion problem.

Several methods aim at producing 10 m S2 bands. Lanaras *et al*. (2018) trained two CNNs to super-resolve the S2 20 m and 60 m bands respectively to 10 m. Their network was

inspired by EDSR (Lim *et al.* 2017) which follows the ResNet architecture (He *et al.* 2016). A residual design was also used by Palsson *et al.* (2018) with the same goal. To super-resolve the 60 m bands their model made use of the 20 m bands in addition to the 10 m bands. Zhu *et al.* (2019) proposed a network that adds channel attention on residual blocks and increased the resolution of S2 20 m bands to 10 m. Gargiulo *et al.* (2019) based on PanNet which uses high-pass filtered inputs (Yang et al., 2017) and a PNN (Masi *et al.* 2016) version that uses a residual skip connection at the end of the network (Scarpa et al., 2018) superresolved S2 20 m bands to 10 m. Wu *et al.* (2020) proposed a parallel residual network (SPRNet) and superresolved S2 60 m and 20 m bands to 10 m. Zhang *et al.* (2020) proposed a GAN that uses residual blocks to enhance the spatial resolution of 20 m (×2) and 60 m (×6) S2 bands by injecting information from the 10 m bands. Finally, Salgueiro *et al.* (2021) proposed a CNN model that takes the 10 m, 20 m, and 60 m S2 bands as input and produces super-resolved 10 m images for the 20 m and 60 m bands. The model uses Residual-in-Residual Dense Blocks (RRDBs) (Wang *et al.* 2019).

Other methods aim at producing higher than 10 m resolutions. Li and Li (2021) used degradation kernel estimation and noise injection to construct a dataset of near-natural LR-HR S2 images and then trained a GAN which composed of an Enhanced Super-resolution (ESR)-GAN-type generator (Wang *et al.* 2019), a PatchGAN-type discriminator and a VGG-19-type feature extractor (Simonyan and Zisserman 2014). Their model produces S2 RGB images with 2.5 m spatial resolution by taking as input the respective 10 m images. In addition, by using two sources of data, *Romero et al.* (2020) trained a model based on ESRGAN and produced 2m RGB-NIR S2 images. The network was first pre-trained with artificially generated WV LR-HR (10 m - 2 m) image pairs and then fine-tuned with S2-WV image pairs.

In this study, concerning the fusion methods adapted from pan-sharpening techniques, the PCA method was selected as it outperformed others in our previous study (Kremezi *et al.* 2021). Regarding the unmixing based methods, the Coupled Nonnegative Matrix Factorization (CNMF) and Lanaras' (alternating unmixing approach) as well as the HySure and the FUSE methods (Bayesian approach) have been selected as they have been proven accurate, reliable, and versatile in terms of their adaptiveness for fusing PAN, MS, and HS data (Kotwal and Chaudhuri, 2013).

Concerning the DL methods, 6 approaches in total were implemented, among which three literature networks (PNN (Masi *et al.* 2016), SRGAN (Ledig *et al.* 2017), RCAN (Zhang *et al.* 2018)) and three that were created for the purpose of the study (Fusion-PNN-Siamese, Fusion-ResNet, Fusion-GAN). The DL literature networks were adjusted to the fusion problem since they originated from either the pan-sharpening (PNN) or the SISR domain (SRGAN, RCAN). PNN was selected because of its simplicity and SRGAN because of its high popularity and the fact that it is the basis of the rest of the super-resolution methods based on GANs. RCAN was implemented because it combines the concept of "attention" with the residual blocks. Inclusion of attention layers in CNNs has shown promising results in many fields.

Fusion-PNN-Siamese, Fusion-ResNet, and Fusion-GAN were designed based on popular DL concepts (parallel branches, residual blocks, adversarial learning, concatenation). Using a low number of trainable parameters, thus producing lightweight networks was also a key requirement.

2.2.2. Spectral unmixing and component substitution methods

The matrix factorization approach (Loncan *et al.* 2015) for HS + MS fusion essentially exploits the spectral unmixing theory. In this case, the Linear Mixed Model (LMM) was used. According to the LMM, each pixel is a linear combination of numerous independent signals and spectral unmixing aims to analyse the composition of these mixed pixels into their endmembers and their corresponding abundances. Two constraints are usually posed on the LMM model: the non-negativity (abundance and endmember values must be positive numbers) and the sum-to-one constraint (the sum of the abundance values must be equal to one). In case that both constraints are posed, the Fully Constrained LMM (FC-LMM) is applied (Equation 1).

$$\mathbf{F} = \mathbf{E} \cdot \mathbf{A} \quad (1)$$

where $\mathbf{F}$ is the fused image, $\mathbf{E} \equiv [e_1, e_2, \dots, e_p]$ is the matrix of all the endmembers present in the image and $\mathbf{A} \equiv [a_1, a_2, \dots, a_{(K \times N)}]$ is the matrix of abundances of each endmember for every pixel of the fused image.

The hyperspectral image $\mathbf{H}$ with low spatial resolution can be expressed as a spatially downsampled $\mathbf{F}$ (Equation 2) while the multispectral image $\mathbf{M}$ can be expressed as a spectrally downsampled $\mathbf{F}$ (Equation 3).

$$\mathbf{H} \approx \mathbf{F} \cdot \mathbf{B} = \mathbf{E} \cdot \mathbf{A} \cdot \mathbf{B} \quad (2)$$

where $\mathbf{B}$ is the downsampling operator in the spatial dimension.

$$\mathbf{M} \approx \mathbf{R} \cdot \mathbf{F} = \mathbf{R} \cdot \mathbf{E} \cdot \mathbf{A} \quad (3)$$

where $\mathbf{R}$ is the spectral response function (SRF) of the $\mathbf{M}$ image.

Both matrices $\mathbf{R}$ and $\mathbf{B}$ can be estimated from the observed data.

In the following analysis, $\mathbf{A}_H$ represents the abundance matrix with the low spatial resolution and $\mathbf{E}_M$ the endmember matrix with the low spectral resolution.

2.2.2.1. Coupled Non-negative Matrix Factorisation (CNMF)

CNMF is a representative HS and MS image fusion method. Using the SPF and SRF matrices that relate the two datasets, the NMF method is alternately applied on the two datasets until convergence (Yokoya *et al.* 2012). The initial matrix $\mathbf{E}$ is estimated by the vertex component analysis (VCA) (Nascimento and Dias, 2005) and then $\mathbf{E}$ and $\mathbf{A}_H$ are alternately optimised using the Lee and Seung's multiplicative rules (Lee and Seung, 1999) (Equation 4).

$$\min\|\mathbf{H} - \mathbf{E}\cdot\mathbf{A}_{\mathrm{H}}\cdot\mathbf{B}\|_F^2 \quad (4)$$

Next, the abundance matrix $\mathbf{A}$ is similarly estimated from the MS data. The initial $\mathbf{E}_{\mathrm{M}}$ is set as the $\mathbf{R}\cdot\mathbf{E}$ and the initial $\mathbf{A}$ is estimated by upsampling $\mathbf{A}_{\mathrm{H}}$ using bilinear interpolation. Then, they are optimised by equation 5.

$$\min\|\mathbf{M} - \mathbf{R}\cdot\mathbf{E}_{\mathrm{M}}\cdot\mathbf{A}\|_F^2 \quad (5)$$

The final fused image is obtained by multiplying the optimised $\mathbf{A}$ and $\mathbf{E}$.

2.2.2.2. Lanaras' method

Similar to CNMF, Lanaras' method alternatively applies spectral unmixing on the two input images to produce the endmember and abundance matrices (Lanaras *et al.* 2015). To optimize the solution to this problem, a projected gradient method based on the Proximal Alternating Linearized Minimisation (PALM) algorithm is used to alternately update the endmember (Equation 5) and the high-resolution abundance (Equation 6) matrices.

The initial matrix E is estimated by the SISAL algorithm and the initial abundance matrix A is estimated by the SUnSAL algorithm which exploits sparse unmixing by variable splitting and augmented Lagrangian (Bioucas-Dias, 2009). The SPF and SRF matrices that relate the two datasets are required to initialize the degraded versions of the endmembers and abundances.

2.2.2.3. Hyperspectral Superresolution (Hysure)

HySure identifies the data fusion as a convex optimization problem which is solved using a form of Vector Total Variation-based (VTV) regularization (Simoes *et al.* 2015). The fusion product is the solution to the following optimisation problem (Equation 6):

$$\min_{\mathbf{A}} \frac{1}{2}\|\mathbf{H} - \mathbf{E}\cdot\mathbf{A}\cdot\mathbf{B}\|_F^2 + \frac{\lambda_m}{2}\|\mathbf{M} - \mathbf{R}\cdot\mathbf{E}\cdot\mathbf{A}\|_F^2 + \lambda_\varphi\varphi(\mathbf{A}) \quad (6)$$

where, λ_m and λ_φ are the relative weights of the different terms and φ is the VTV regularizer (Equation 7).

$$\varphi(\mathbf{A}) = \sum_{i=1}^{n}\sqrt{\sum_{k=1}^{m_\lambda}\left\{\left[(AD_h)_j^k\right]^2 + \left[(AD_v)_j^k\right]^2\right\}} \quad (7)$$

where m_λ is the number of the HS image bands and n is the number of pixels. Moreover, $(AD)_j^k$ denotes the element in the kth row and jth column of the product of matrix $\mathbf{A}$ by matrices $\mathbf{D}_{\mathrm{h}}$ and $\mathbf{D}_{\mathrm{v}}$ respectively, which compute the horizontal and vertical discrete differences of an image with periodic boundary conditions.

The above problem is solved by following an Alternating Direction Method of Multipliers (ADMM) approach by using the Split Augmented Lagrangian Shrinkage Algorithm (SALSA) (Afonso *et al.* 2011).

2.2.2.4. Fast fUsion based on Sylvester Equation (FUSE)

FUSE solves a Sylvester equation associated with the fusion problem derived from Equations 2-3 (Wei *et al.* 2015b). It aims at solving the maximization problem of the likelihoods obtained from the forward observation models. The computational performance of the algorithm is improved by implementing a closed-form solution to the Sylvester equation. Similarly to the aforementioned methods, FUSE requires prior knowledge of the SPF and SRF matrices. Moreover, the algorithm is applied on the principal component subspace and the naïve Gaussian priors are used as a priori information.

2.2.2.5. PCA based method

In addition to the unmixing-based methods, a method adapted from pan-sharpening techniques was implemented. It is a Component Substitution (CS) approach, where a component of the HS image is substituted with a component of the MS image. This method exploits the Principal Component Analysis (PCA). Both images are projected into the principal component space and subsequently, the HS component that contains the spatial information is substituted with the respective MS component. Finally, the fused data are projected back to the original space.

2.2.2.6. Optimal band estimation

The non-DL methods were applied using various WV band combinations (Table 1) to determine the optimal number of bands and spectral range for downscaling S2 images. Due to the higher computational demand of the DL approaches, the investigation of the optimal band combination was carried out only on the non-DL methods.

Table 1
WV band combinations

Combination	WV Bands
all9	All 9 (VNIR +SWIR)
all7	All7 (VNIR)
234	Blue – Green – Red
2346	Blue – Green – Red – NIR1
2347	Blue – Green – Red – NIR2
1234	Coastal - Blue – Green – Red
2348	Blue – Green – Red – NIR1 – SWIR3
23469	Blue – Green – Red – NIR1 – SWIR6
23478	Blue – Green – Red – NIR2 – SWIR3
23479	Blue – Green – Red – NIR2 – SWIR6

2.2.3 Deep learning (DL) methods

For the fusion of the 13 bands of the S2 image with the bands of the WV-3 image acquired on 7/6/2018, 6 DL approaches were implemented. The first two networks (Fusion-PNN, Fusion-PNN-Siamese) were created by modifying the CNN architecture proposed by Masi *et al.* (2016) (PNN). Since PNN was designed for pan-sharpening, which can be considered a subset of the data fusion problem, the input layer of the CNN was modified to suit the fusion of the S2 and WV-3 images. The third network (Fusion-ResNet)

was constructed by combining residual blocks and layer concatenation. The above popular concepts were first proposed in the ResNet and UNet (Ronneberger *et al.* 2015) architectures respectively. The fourth network (Fusion-GAN) used a generative adversarial approach and the generator architecture was based on UNet. Finally, the last two DL approaches were the implementations of SRGAN (super-resolution GAN) and RCAN (residual channel attention network). Since SRGAN and RCAN were designed to solve the super-resolution problem, the input layer was modified to suit the fusion problem.

Fusion-PNN was trained with two different band configurations: i) 8 VNIR and ii) 8 VNIR + 2 SWIR, while the rest of the methods were trained only with the first band configuration. The selected SWIR WV-3 bands correspond to 1640-1680 nm and 2185-2225 nm. The selection was based on the fact that these bands are the spectrally closest to the last two S2 SWIR bands (1613 nm, 2202 nm).

2.2.3.1. Architecture – Activation functions

All DL approaches except for Fusion-PNN-Siamese used the Early Fusion (EF) method, i.e. the input of the network was created by concatenating the S2 and WV-3 bands. The sequence of the bands matched the sequence of the corresponding wavelengths. In Fusion-PNN-Siamese, the S2 and WV-3 bands were fed as input to two different branches. For the creation of the Siamese network, the outputs of the second convolutional layer of the two branches were concatenated and then fed to a third convolutional layer. In both PNN versions (Figures 3-4), the suggestions of *Masi et al.* (2016) were followed for the number of feature maps, the activation function and the size of kernels in the first and second convolutional layers (64/ReLU/9×9, 32/ReLU/5×5), as well as the kernel size of the third convolutional layer (5×5 px). The sigmoid activation function (Nwankpa *et al.* 2018) (Equation 8) was applied in the output layer.

The equation of the ReLU activation function (Agarap, 2018) is presented in Equation 9.

$$\varphi(x) = \frac{1}{1+e^{-x}} \quad (8)$$

$$\varphi(x) = \max(0, x) \quad (9)$$

where x is the output produced by the convolution operation.

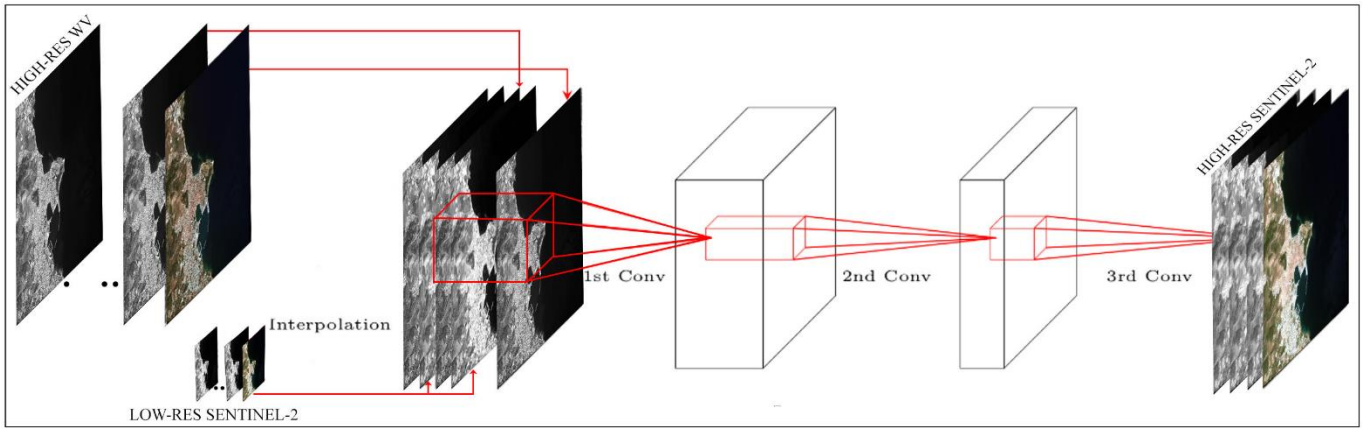

Fig. 3. Architecture of Fusion-PNN.

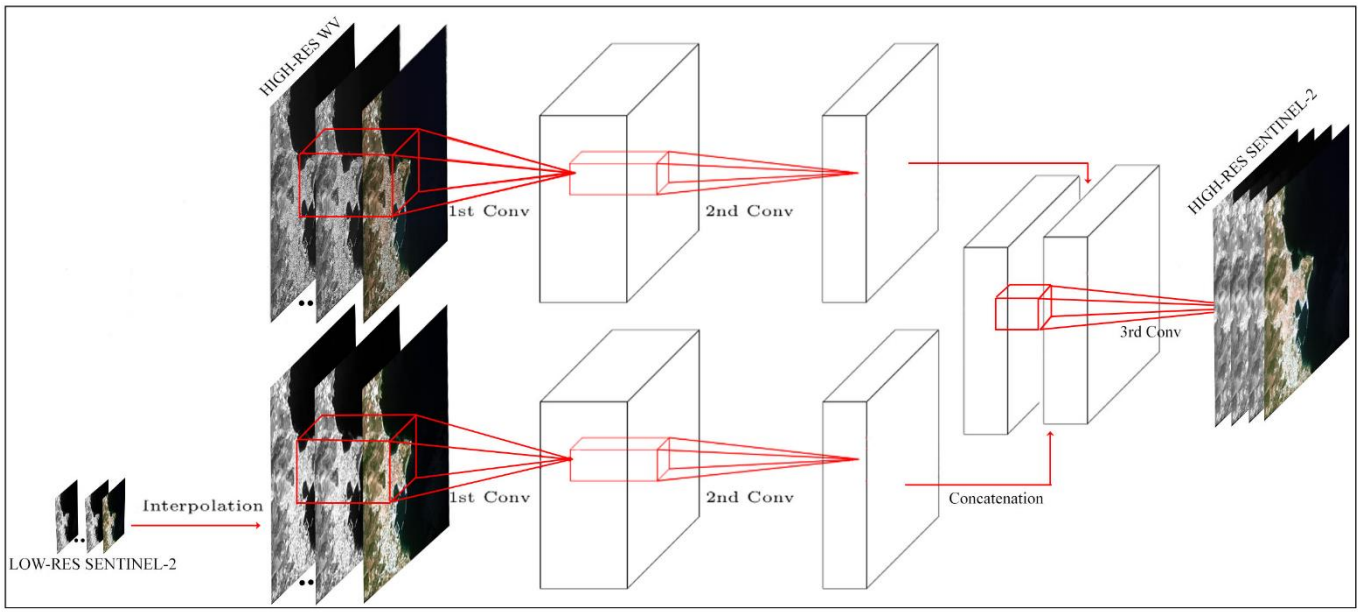

Fig. 4. Architecture of Fusion-Siamese-PNN.

The architecture of Fusion-ResNet is shown in Figure 5. The encoding part of the network was initialized with a convolutional layer and it was also composed of 4 identity residual blocks and two convolutional residual blocks. The decoding part included two transposed convolutional layers and a convolutional layer towards the end. Skip connections through concatenation were applied between the encoder and the decoder. Further details are shown in Figure 5 (e.g. activation-batch normalization layers, convolution hyperparameters, number of feature maps etc.). It is noted that input and output images shown in Figure 5 correspond to the inference resolution.

The architecture of Fusion-GAN is shown in Figure 6. The network was composed of a generator and a discriminator. The generator encoder consisted of 5 convolutional layers and the decoder of two transposed convolutional layers and a convolutional layer towards the end. Skip connections through concatenation were applied between the encoder and the decoder. Concerning the discriminator, it consisted of 4 convolutional layers. Further details are shown in Figure 6. It is noted that input images shown in Figure 6 correspond to the training resolution.

SRGAN was implemented with the hyperparameters suggested by Ledig *et al.* (2017). The upsampling layers towards the end of the network were removed because our input x-y size matched the output size. Among others, the generator contained 16 residual blocks and the discriminator 8 convolutional layers where the number of feature maps gradually increased (64→512). Further details are shown in the respective paper.

RCAN uses the residual in residual structure and incorporates a channel attention module. The settings proposed by Zhang *et al.* (2018) were followed for its implementation with the difference that we selected 5 residual groups and 10 residual channel attention modules instead of 10 and 20 to ensure faster training time. Further details are shown in the respective paper.

2.2.3.2. Pre-processing – Training - Inference

Given that ground-truth data at the high-resolution (4 m S2 image) are not available, the DL approaches were trained based on the assumption that the spatial details are self-similar and scale-invariant as considered in previous works (e.g. Lanaras *et al.* 2018; Zhu *et al.* 2019; Wu *et al.* 2020b; Salgueiro *et al.* 2021). Thus, we assumed that super-resolving from 20 m to 4 m can be learned from super-resolving at a reduced resolution where ground-truth data are available (Wald's protocol). In more detail, during training the inputs of the CNN were: i) the WV-3 bands downsampled to 20 m spatial resolution and ii) the Sentinel-2 bands downsampled to 100 m and then upsampled to 20 m. Thus, the spatial resolution

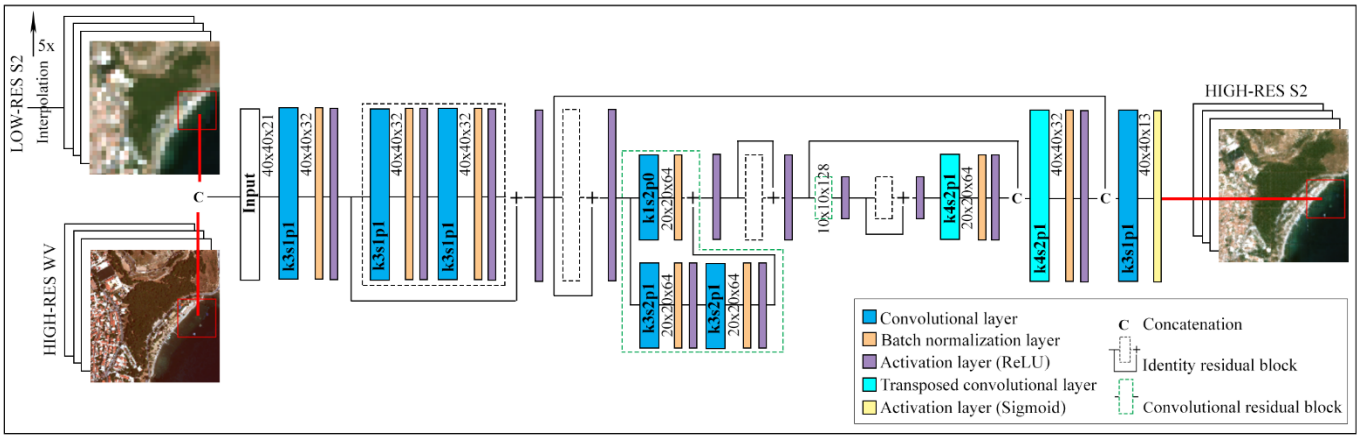

Fig. 5. Architecture of Fusion-ResNet.

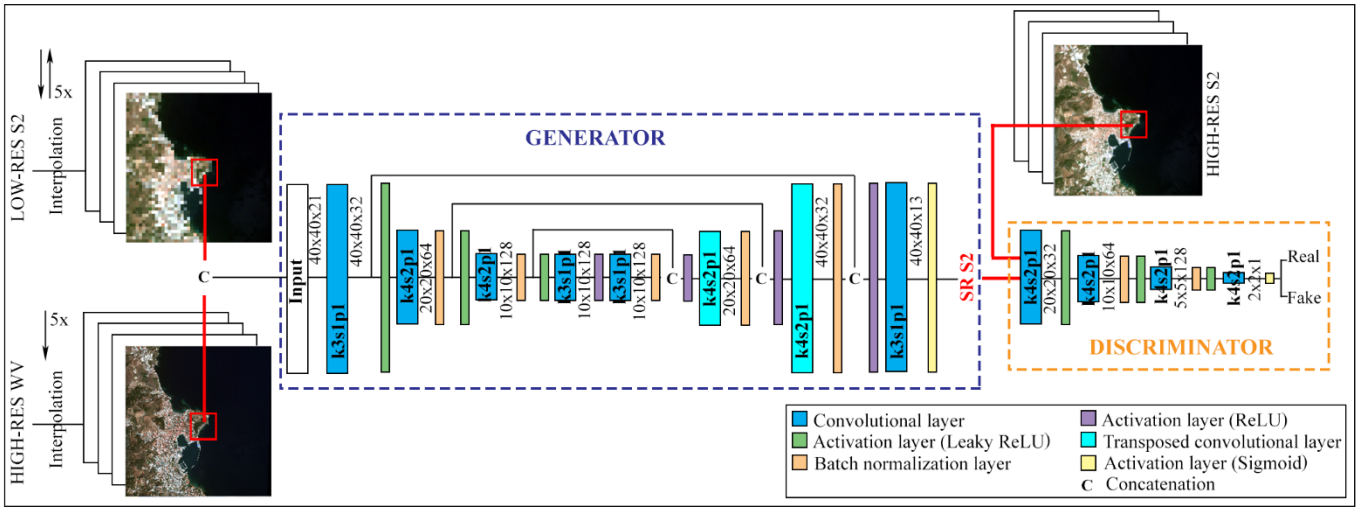

Fig. 6. Architecture of Fusion-GAN.

ratio between the WV-3 and S2 bands was 1/5. During the inference stage, the fused output image (spatial resolution: 4 m) was created by feeding the network with i) the WV-3 bands with 4 m spatial resolution and ii) the S2 bands with 20 m spatial resolution. Nearest neighbour interpolation was used during all resampling operations which led to the same x-y size between the WV-3 and the S2 bands. In addition, for each band, 1% of the histogram values (left and right) were clipped to prevent lower CNN performance due to sparse extreme values.

During training, the Adaptive moment estimation (Adam) (Kingma and Ba, 2014) method was used to update the weights in the backpropagation process. Concerning the loss function, the mean squared error (MSE) was selected as the Loss function for Fusion-PNN, Fusion-PNN-Siamese, and Fusion-ResNet (Equation 10) and the L1 loss (Equation 11) for RCAN. In the adversarial training (Fusion-GAN, SRGAN), the generator and the discriminator were trained in an alternating way according to the adversarial loss function shown in Equation 11 (Goodfellow *et al.* 2014). In addition, since the problem belongs in the super-resolution domain, a content loss was added to the generator loss to significantly increase the performance (Ledig *et al.* 2017) (Equation 13 (Fusion-GAN), Equation 14 (SRGAN)). It is noted that the originally proposed VGG loss for SRGAN was substituted with the pixel-wise MSE loss because our output contained 13 bands instead of three, thus the VGG Imagenet (Fei-Fei *et al.* 2010) pre-trained weights could not be used.

Table 2
Training details

Training details	Fusion-PNN_VNIR	Fusion-PNN_VNIR+SWIR	Fusion-PNN-Siamese	Fusion-ResNet	Fusion-GAN	SRGAN	RCAN
Loss metric	0.0007 (MSE)	0.0006 (MSE)	0.0006 (MSE)	0.0002 (MSE)	0.0101 (L1)	0.0017 (MSE)	0.0014 (L1)
Epochs	4000	4000	4000	800	800	500	500
Batch size	128	128	128	128	128	16	16
Patch size	9x9	9x9	9x9	40x40	40x40	40x40	40x40
Number of patches	48600	48600	48600	6400	6400	6400	6400
Trainable params	170,573	180,941	232,269	906,029	974,542	10,812,831	3,999,957
Library	Keras/TF	Keras/TF	Keras/TF	Pytorch	Pytorch	Pytorch	Pytorch
Training time (h)	2.0	2.2	3.0	1.1	0.9	7.5	10.0

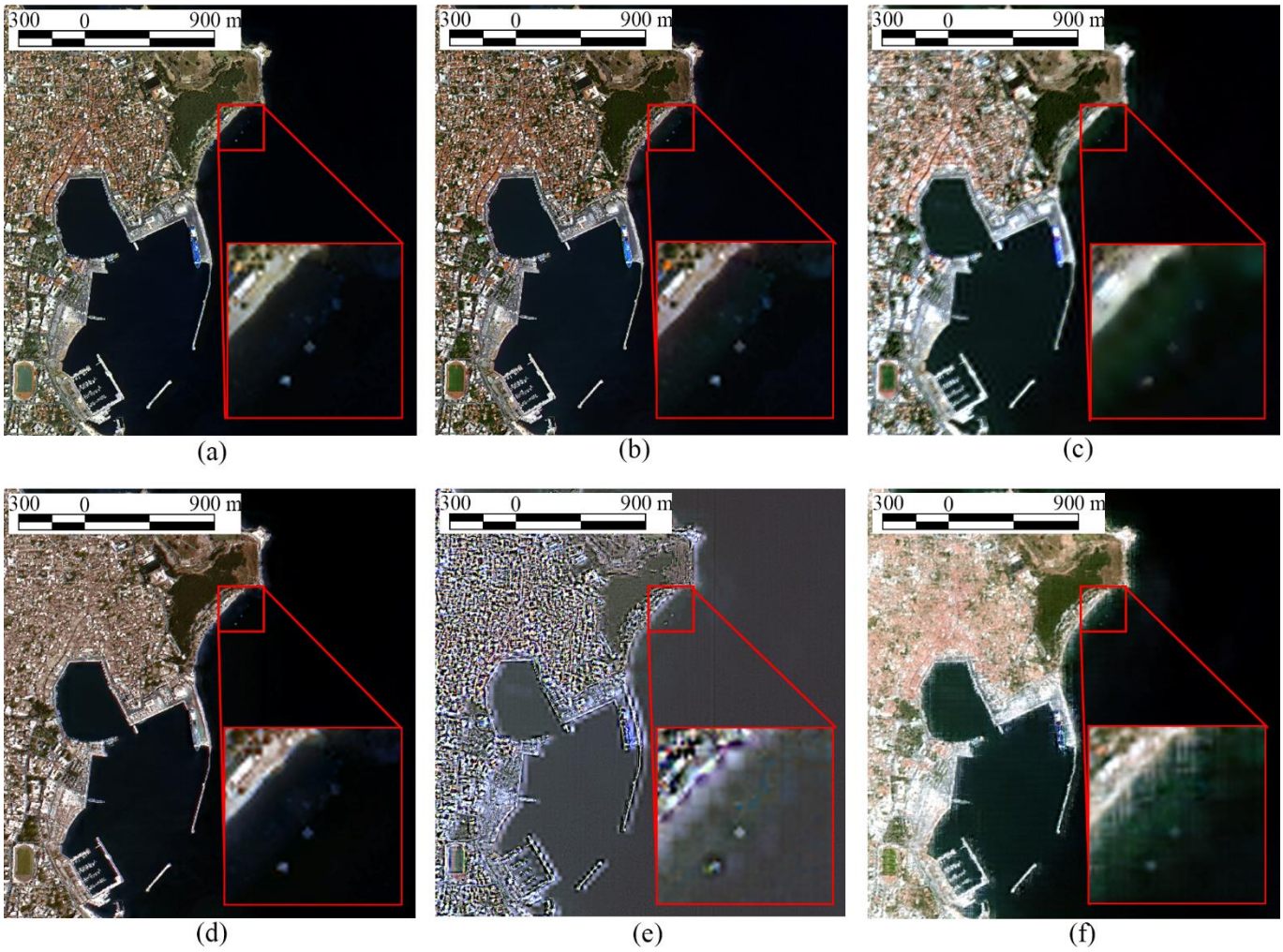

Fig. 7. Fusion results for the S2 and WV-3 images acquired on 07/06/2018 (zoomed-in view of the target placement in red window) (Natural colours). (a) CNMF 2347, b) HySure all9. (c) PCA all9. (d) Lanaras' 1234. (e) FUSE 2348. (f) Fusion-PNN_VNIR

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \quad (10)$$

$$\mathrm{L1} = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| \quad (11)$$

$$min_G max_D \begin{bmatrix} \mathbb{E}_{(\mathrm{HR}\sim P(\mathrm{HR}))}[logD(\mathrm{HR})] + \\ \mathbb{E}_{(\mathrm{LR}\sim P(\mathrm{LR}))}[\log(1 - D(G(\mathrm{LR})] \end{bmatrix} \quad (12)$$

$$L_{\mathrm{G\_Fusion-GAN}} = -\log\left(D\left(G(\mathrm{LR})\right) + 10^2\mathrm{L1}\right. \quad (13)$$

$$L_{\mathrm{G\_SRGAN}} = -10^{-3}\log\left(D\left(G(\mathrm{LR})\right) + \mathrm{MSE}\right. \quad (14)$$

where Y_i are the observed spectral values, $\hat{Y}_i$ are the predicted spectral values, and n is the number of pixels in a patch.

Training details are given in Table 2. All models were trained on an NVIDIA 1070 Ti GPU. The x-y size of the images during training was 229×234 px and during inference was 1145×1170 px.

3. Image fusion results and evaluation

3.1. Fusion of S2 and WV-3 (07/06/2018 experiment)

A CS, 4 unmixing-based, and 6 DL approaches for image fusion were evaluated for the 07/06/2018 experiment. The fused images of the various methods have 4 m spatial resolution and 13 bands. Natural colour composites are presented in Figures 7-8. For the CS and unmixing methods, the composites for the band combination (Table 1) that produced the best fusion results are presented. CNMF and HySure outperformed all the other methods in terms of spatial information. Their results present clear edges without any blurring or remaining artifacts and all plastic targets are discernible. Less satisfactory results were produced by the DL approaches. Fusion-ResNet and Fusion-GAN show less noisy outputs compared to the other DL methods. The lower performance could be explained by training in the reduced scale (lack of relevant high-frequency information) because of the unavailability of ground-truth data and the high ratio (1/5). It is noted that contrary to the low performance in the high resolution, all DL methods produced high MSE or L1 scores in the reduced resolution (Table 2).

The conclusions reached by the visual interpretation of Figures 7-8 are confirmed by image quality metrics (i.e PSNR, ERGAS, RMSE, and SSIM (Jagalingam and Hegde, 2015)) and indeed, CNMF_2347 presents the best results according to 3 out of 4 metrics (Table 3).

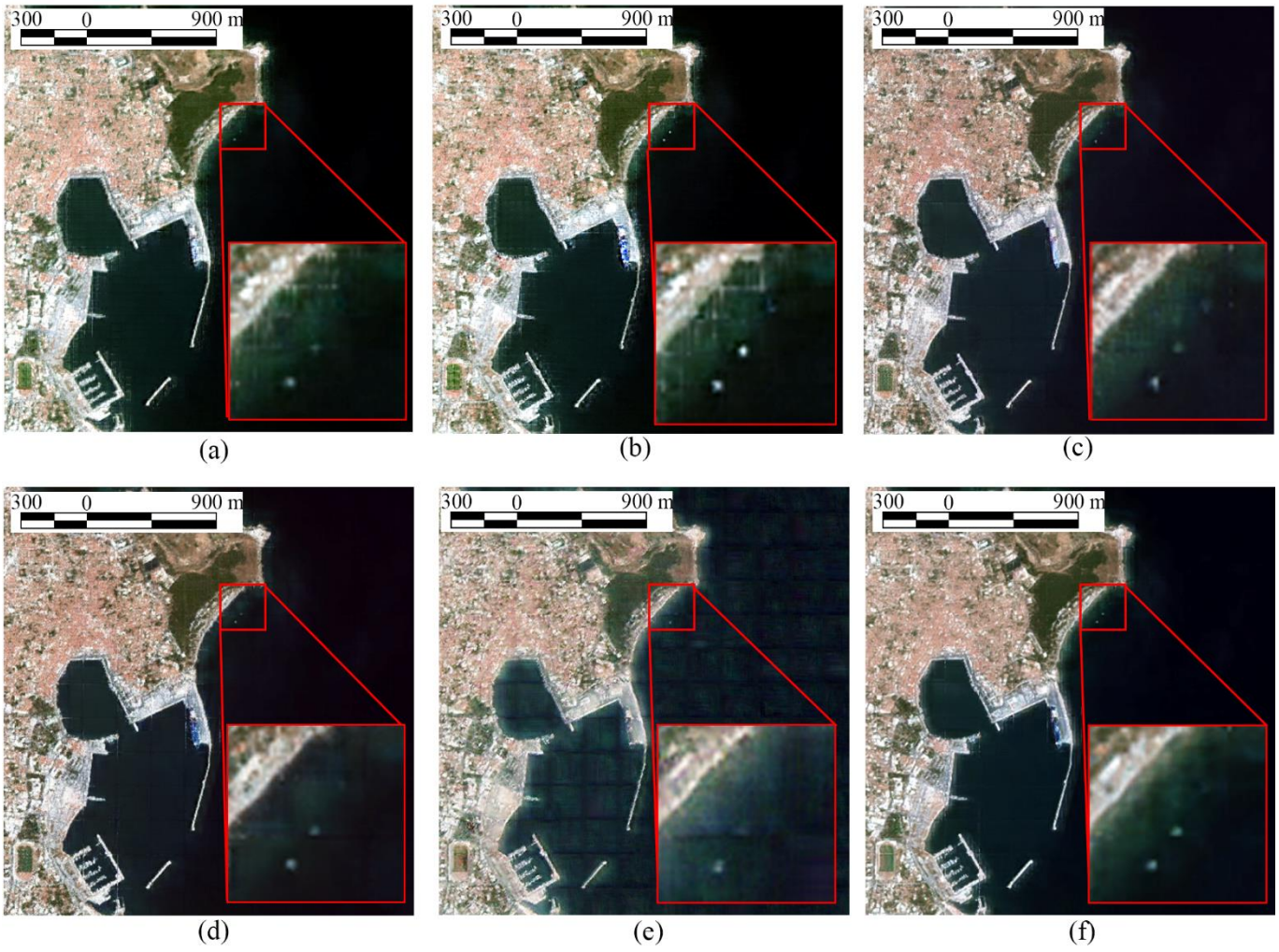

Fig. 8. Fusion results for the S2 and WV-3 images acquired on 07/06/2018 (zoomed-in view of the target placement in red window) (Natural colours). (a) Fusion-PNN_VNIR+SWIR, b) Fusion-PNN_Siamese, (c) Fusion-ResNet, (d) Fusion-GAN, (e) SRGAN, (f) RCAN

Table 3
Quality metrics between fusion results and WV reference images

Fusion Method	PSNR	ERGAS	RMSE	SSIM
FUSE_2348	5.89	64.38	331.57	0.192
FUSE_all7	8.71	46.69	239.69	0.067
Lanaras'_1234	9.01	44.35	234.55	0.300
Lanaras'_2347	8.11	49.20	260.48	0.321
CNMF_2347	**27.24**	**5.45**	**29.06**	0.490
CNMF_all7	25.68	6.54	34.94	**0.597**
HySure_all7	19.71	12.98	67.64	0.559
HySure_all9	21.25	10.89	56.65	0.571
PCA_23478	15.24	21.64	114.27	0.028
PCA_all9	15.46	21.10	111.37	0.028
Fusion-PNN_VNIR+SWIR	4.65	73.30	388.60	0.069
Fusion-PNN_VNIR	4.62	73.49	389.47	0.068
Fusion-GAN	5.21	68.71	365.10	0.072
SRGAN	4.45	74.97	396.85	0.052
RCAN	5.78	64.50	343.94	0.075
Fusion-ResNet	5.05	69.99	371.39	0.063
Fusion-PNN-Siamese	4.88	71.39	378.85	0.068

Spectral signatures were also evaluated (Figures 9-12). In these Figures, the spectral signatures of the "plastic bottles", "fishing nets", "plastic bags" targets and of a random water pixel respectively, are shown for all fused results as well as for the original S2 and WV3 (resampled to 20 m) images. In Table 4, similarity between these spectra from the fusion results and the reference S2 images are examined using the measure of spectral angle distance (SAD) and Correlation Coefficient (CC) (Jagalingam and Hegde, 2015). It can be seen that all DL methods except for SRGAN outperformed the non-DL methods in preserving the S2 spectral characteristics. Among the DL methods, best performance was shown by Fusion-ResNet and Fusion-GAN which are lighter networks than RCAN and SRGAN as it can be seen from the number of trainable parameters shown in Table 2. CNMF_2347 yields also a good spectral performance.

Overall, water pixels present the lowest SAD values. This is attributed to the fact that the pixel coverage of the targets in the 4 m /2 m fused image is higher than the pixel coverage in the original 20 m S2 image. SAD values between water pixels are low because in both images the water pixels are pure. SAD values between plastic pixels are higher because the S2 plastic pixel is mixed with water, while the WV2/3 plastic pixel is pure plastic (Figure 13). As shown in Figure 14, the resulting pure plastic pixels in the fused image enable inter- and intraclass separability.

The WV-3 reference spectra mostly differ from the S2 reference spectra at around 1.0 – 1.5 μm (Figures 9-12). This could be explained by observing the lack of overlap in the respective SRFs of the sensors in this spectral region (Figure 15). From our experiments we did not reach safe conclusions about correlations between the performance of the fusion methods and the similarity of SRFs. This would require further analysis which is out of the scope of this study.

3.2. Fusion of S2 and WV-2 (21/06/2021 experiment)

The method with the optimal results (CNMF_2347) on the 07/06/2018 experiment, was applied on the S2 and WV-2 data in the second experiment. The fused image has 13 bands and 2 m spatial resolution. Figure 18 presents the resulting image, where it is observed that all targets are visible in the WV-2 image except for the 0.6 m PET target, which was not visible in the original WV-2 image either.

In Figures 16 and 17 it can be seen that spectral information is retained in the fused output. After image fusion, there is high separability with water spectra maintaining low reflectance values while plastic spectra present quite higher reflectance values. This behaviour concurs with the results of the first experiment. For the image of the 21/06/2021 experiment, only signatures from the big circular targets were extracted since the other targets are not visible in the original S2 image. Considering the separability between the various plastic materials (Figure 17), similarly to Figure 14, it can be seen that it is possible. Their identification though needs to be further examined and it will be the subject of future work.

3.3. Marine litter indexes

3.3.1 Related work

Most marine indexes that have been proposed in the literature have focused in detecting vegetation pixels (e.g. sargassum). The most popular indexes are the Normalized Difference Vegetation Index (NDVI, (Rouse *et al.* 1973)), the

Table 4
Spectral angle distances and correlation coefficients between fusion results and S2 reference spectra

	SAD						CC					
Fusion Method	Plastic Bottles	Fishing Nets	Plastic Bags	Water Pixel 1	Water Pixel 2	Water Pixel 3	Plastic Bottles	Fishing Nets	Plastic Bags	Water Pixel 1	Water Pixel 2	Water Pixel 3
FUSE_2348	0.37098	0.37687	0.28072	0.09187	0.03267	0.04300	0.76207	0.77718	0.89632	0.99360	0.99874	0.99888
FUSE_all7	0.36224	0.44848	0.41681	0.13655	0.03159	0.36566	0.77258	0.68648	0.83925	0.97591	0.99866	0.84242
Lanaras'_1234	0.24236	0.16581	0.20032	0.03918	0.02284	0.06292	0.92348	0.95368	0.96610	0.99951	0.99944	0.99792
Lanaras'_2347	0.33472	0.19233	0.19853	0.03232	0.01730	0.04230	0.81112	0.95290	0.94393	0.99937	0.99960	0.99896
CNMF_2347	0.24073	0.18939	0.15737	0.03101	0.02657	0.03562	0.92500	0.95776	0.96313	0.99885	0.99919	0.99950
CNMF_all7	0.37138	0.28599	0.14775	0.03040	0.03516	0.01809	0.77054	0.90493	0.97559	0.99910	0.99912	0.99979
HySure_all7	0.35376	0.24772	0.19551	0.08176	0.07707	0.12098	0.79265	0.92077	0.95011	0.99517	0.99613	0.99220
HySure_all9	0.35525	0.23990	0.18414	0.08190	0.07770	0.12228	0.78959	0.92578	0.96016	0.99522	0.99609	0.99202
PCA_23478	0.42454	0.23186	0.17579	0.02572	0.03416	0.08236	0.66803	0.94265	0.97053	0.99953	0.99921	0.99701
PCA_all9	0.41877	0.21243	0.18778	0.02630	0.02649	0.07233	0.67928	0.95398	0.96542	0.99951	0.99950	0.99773
Fusion-PNN_VNIR+SWIR	0.25883	0.10933	0.08900	0.00928	**0.00909**	0.00962	0.91513	0.99088	0.99113	0.99990	**0.99991**	0.99992
Fusion-PNN_VNIR	0.09894	0.07254	0.13386	0.00912	0.01276	**0.00511**	0.99165	0.99419	0.98222	0.99989	0.99979	**0.99997**
Fusion-GAN	**0.03372**	**0.05548**	**0.05377**	0.01514	0.01417	0.00944	**0.99835**	**0.99599**	**0.99621**	0.99987	0.99981	0.99994
SRGAN	0.24277	0.21780	0.20430	0.21516	0.18271	0.19169	0.90649	0.92685	0.93819	0.93664	0.95560	0.95782
RCAN	0.09238	0.08684	0.09470	0.01033	0.01118	0.00878	0.99122	0.99066	0.98900	**0.99991**	0.99988	0.99991
Fusion-ResNet	0.05893	0.07196	0.09577	0.00982	0.01300	0.00918	0.99444	0.99382	0.98990	0.99987	0.99979	0.99991
Fusion-PNN-Siamese	0.25876	0.27071	0.10190	**0.00843**	0.00921	0.00727	0.90122	0.90228	0.98835	0.99990	0.99989	0.99994

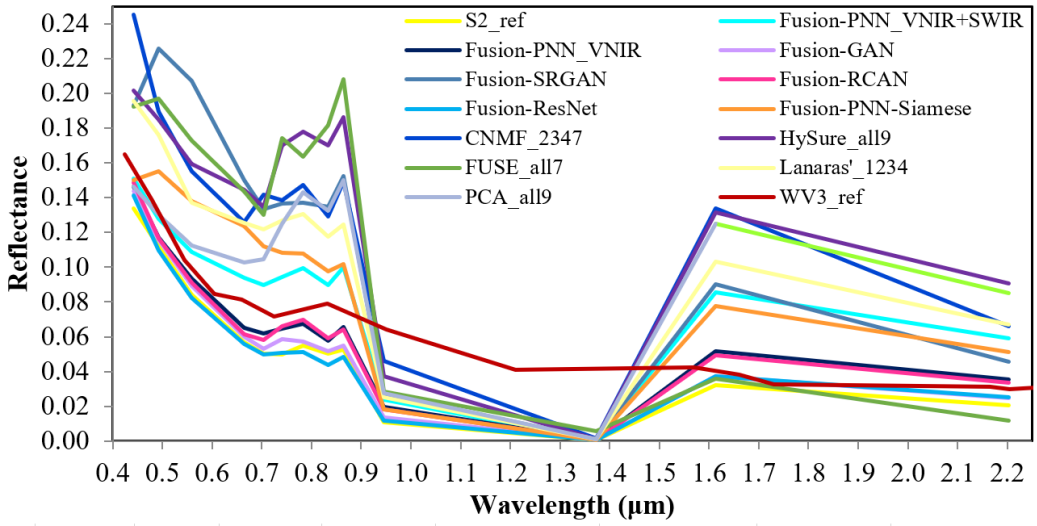

Fig. 9. Spectral signatures of plastic bottles from all fusion results. S2 and WV-3 reference spectra are also included.

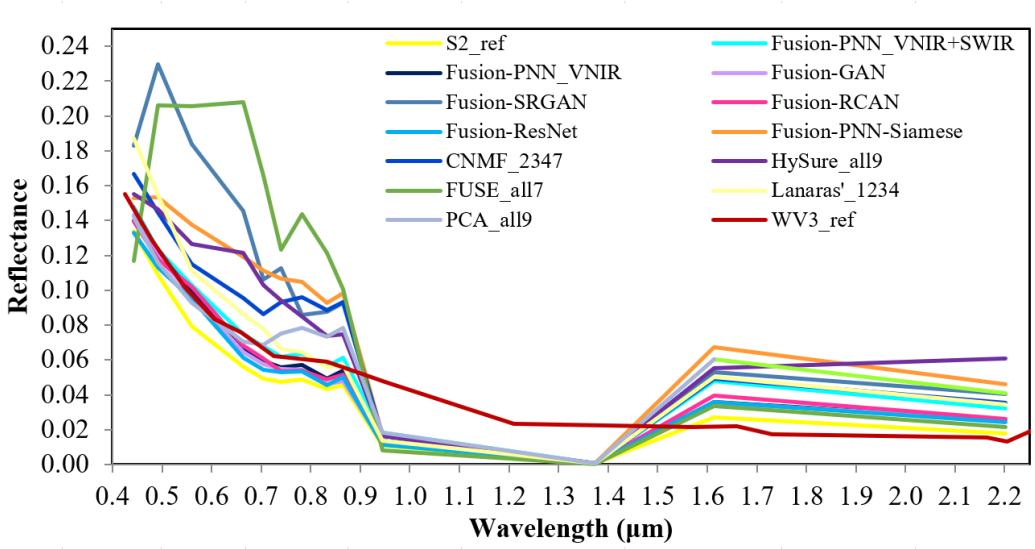

Fig. 10. Spectral signatures of fishing nets from all fusion results. S2 and WV-3 reference spectra are also included.

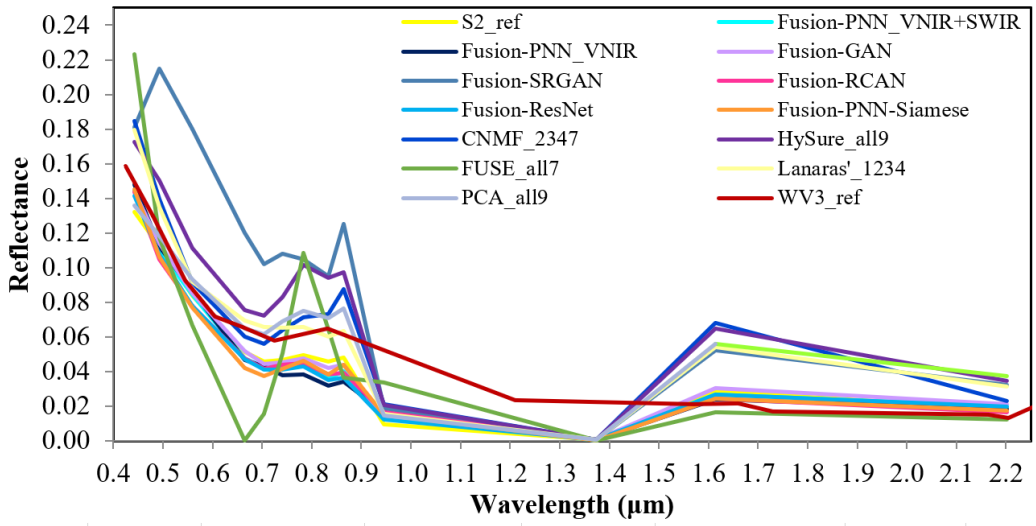

Fig. 11. Spectral signatures of plastic bags from all fusion results. S2 and WV-3 reference spectra are also included.

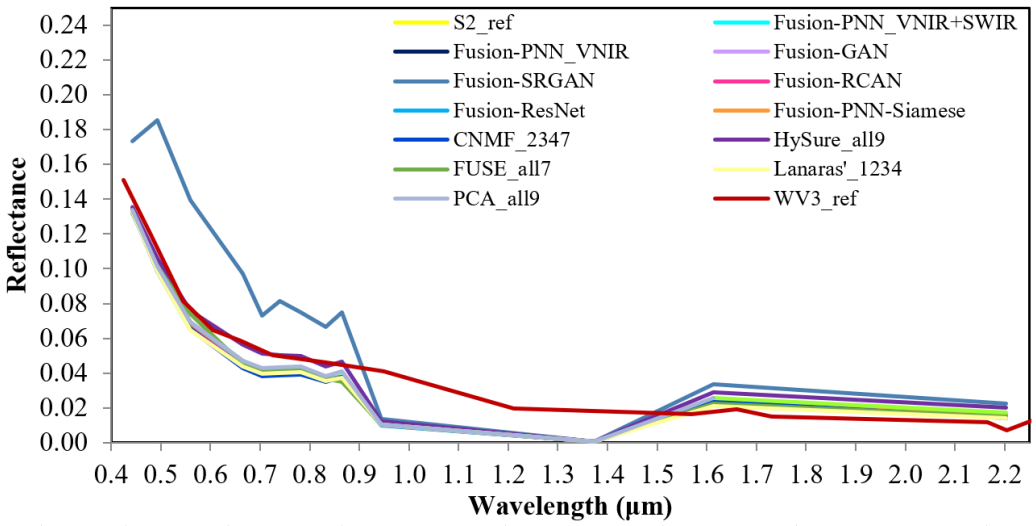

Fig. 12. Spectral signatures of a random water pixel from all fusion results. S2 and WV-3 reference spectra are also included.

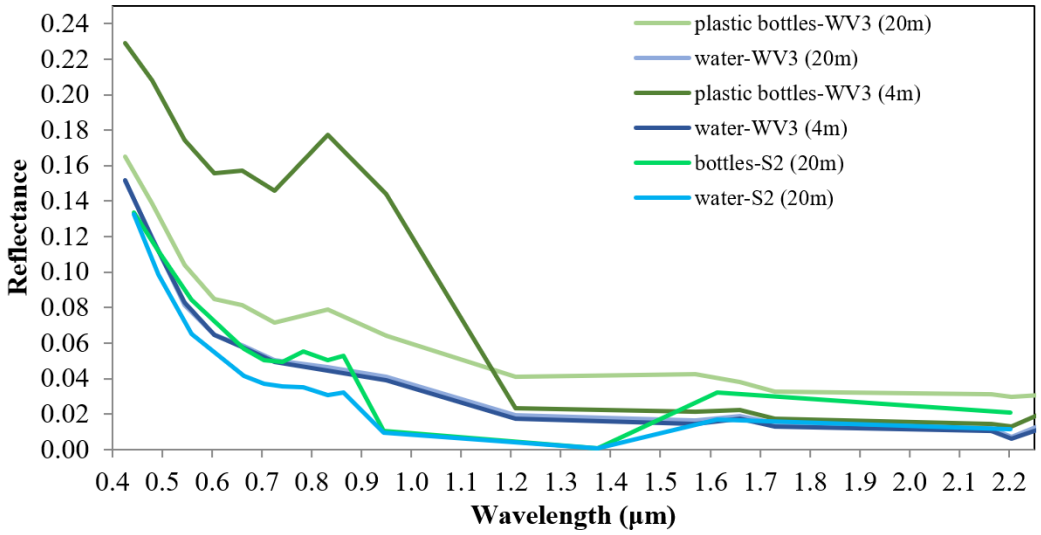

Fig. 13. Spectral signatures of the plastic bottles target and a water pixel from the S2 and WV-3 reference images.

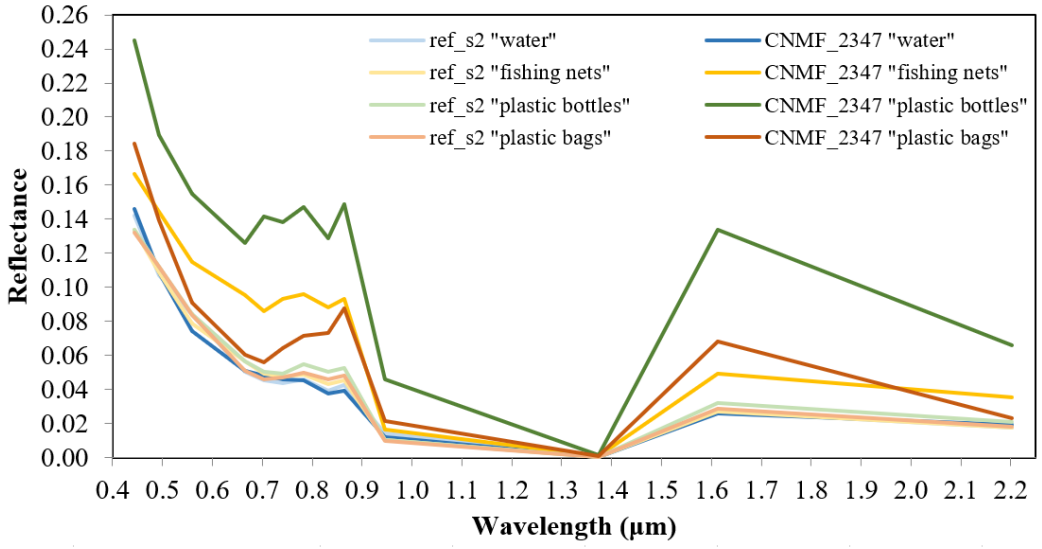

Fig. 14. Comparison of water and plastic target spectra before and after S2 and WV-3 fusion with CNMF method and 2347 band combination (07/06/2018 experiment).

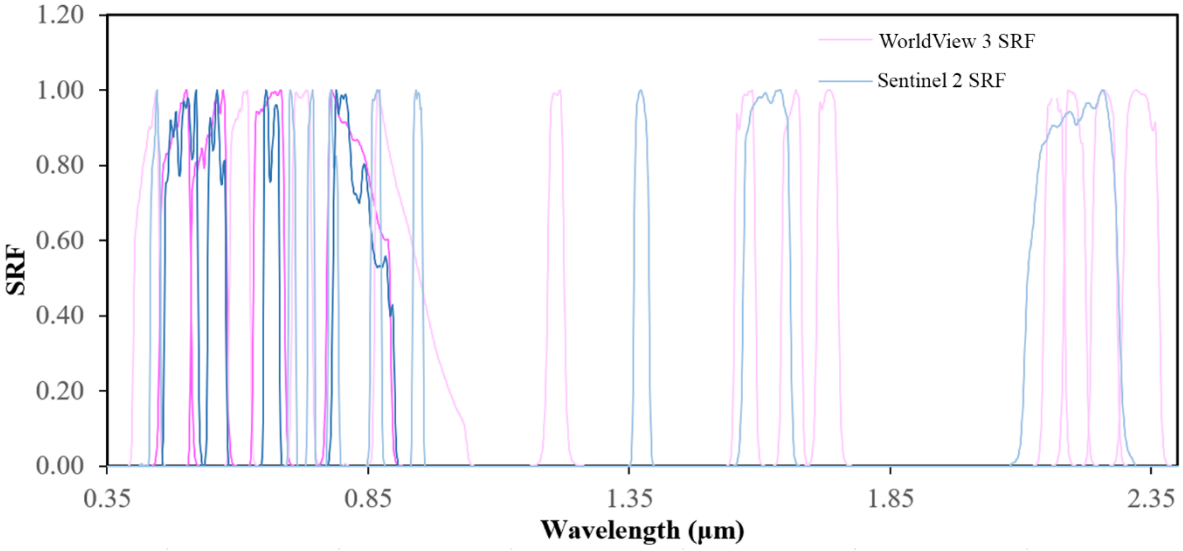

Fig. 15. Spectral Response Functions of the S2 and WV-3 sensors.

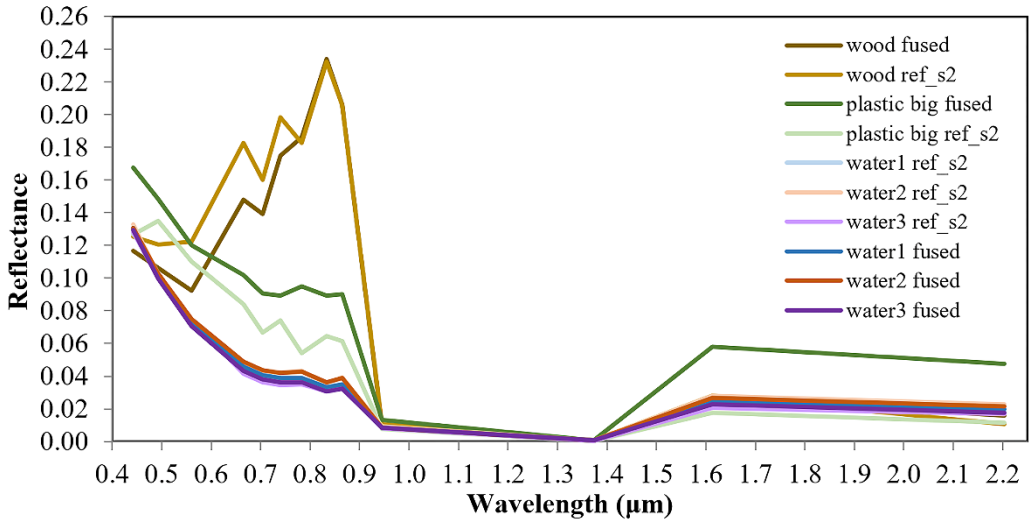

Fig. 16. Comparison of water, large plastic target and wood target spectra before and after S2 and WV-2 fusion with CNMF method and 2347 band combination (21/06/2021 experiment).

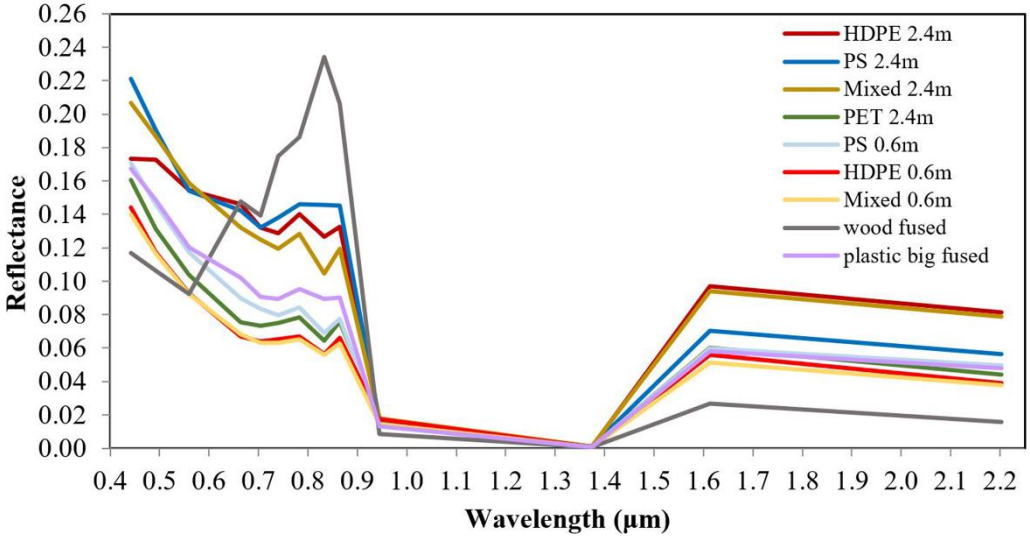

Fig. 17. All plastic target spectra and wood spectra after S2 and WV-2 fusion with the CNMF method and 2347 band combination for the 21/06/2021 image.

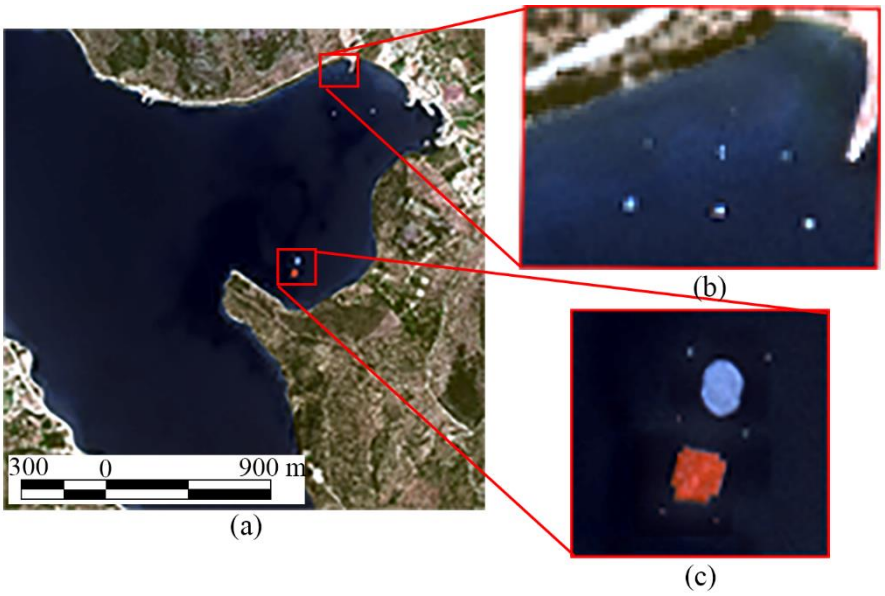

Fig. 18. CNMF_2347 fusion result for the S2 and WV-2 images acquired on 21/06/2021. (a) full image, (b) zoomed-in view of the square targets' placement and (c) zoomed-in view of the circular targets' placement (Natural colours).

Hydrocarbon Index (HI, Kühn *et al.* 2004), and the Floating Algae Index (FAI) which has been developed for Landsat, Medium Resolution Imaging Spectrometer (MERIS), and Moderate Resolution Imaging Spectroradiometer (MODIS) (Hu *et al.* 2015; Hu 2009; Wang and Hu 2016). The indexes use either band subtraction or ratio operations, the former being considered less sensitive to environmental conditions (Hu 2009; Hu *et al.* 2012).

Several researchers have studied the performance of the aforementioned indexes in discriminating vegetation from plastic based on the fact that vegetation causes elevated Red Edge reflectance in contrast to marine debris that causes only elevated NIR values (Hu et al., 2015). Biermann *et al.* (2020) and Kikaki *et al.* (2020) also observed that vegetation shows higher values in the visible range compared to marine debris. Garaba and Dierssen (2018) applied HI and observed an overlap between vegetation and landfill plastics because natural hydrocarbons have similar absorption features to synthetic hydrocarbons and assumed that NDVI or FAI could be more efficient in differentiating vegetation from plastics. In addition, Hu *et al.* (2015) concluded that FAI could potentially differentiate vegetation from garbage given sufficient pixel coverage.

Other researchers have focused more in the detection of marine debris. Papageorgiou (2019) developed the S2-based index by observing reverse slope between S2 plastic and water spectra in the range 780-833 nm and 833-865 nm. In addition, Biermann *et al.* (2020) developed FDI based on FAI to identify floating debris. In FDI the red band was substituted with the Red Edge band. However, they showed that NDVI displayed higher performance in separating plastic pixels from other categories (timber, foam, pumice). FDI combined with NDVI and three spectral bands (Red Edge, NIR, SWIR) was proposed as input to a Naïve-Bayes classifier which achieved 86% accuracy. Both above mentioned studies exploited S2 spectra from the experiment conducted in 07/06/2018 (Topouzelis *et al.* 2019).

3.3.2 Evaluation of the marine litter indexes

All the fused images have very high spatial resolution and three bands in the SWIR region, consequently, they can be exploited by various marine litter indexes which rely on discriminative features of plastic and water spectra. Five indexes were applied on the CNMF_2347 fused images: HI, the S2-based index, FDI, NDVI and FAI (Equations 15-20). HI was tested on three different band combinations based on absorption features observed on plastic spectra: 1) 560, 665, and 783 nm, 2) 560, 740, and 865 nm, and 3) 783, 833, and 865 nm. The best result was produced by the 3rd combination.

$$NDVI = \frac{R_{NIR}-R_{RED}}{R_{NIR}+R_{RED}} \quad (15)$$

$$FAI = R_{NIR} - R_{RED} - (R_{SWIR} - R_{RED})\frac{\lambda_{NIR}-\lambda_{RED}}{\lambda_{SWIR}-\lambda_{RED}} \quad (16)$$

$$FDI = R_{NIR} - R_{RE2} - (R_{SWIR} - R_{RE2})\frac{\lambda_{NIR}-\lambda_{RE2}}{\lambda_{SWIR}-\lambda_{RE2}} \quad (17)$$

$$S2-based = \frac{(R_{RE3}-R_{NIR})+(R_{NNIR}-R_{NIR})}{R_{RE3}+R_{NNIR}} \quad (18)$$

$$HI = (\lambda_B - \lambda_A)\frac{R_C-R_A}{\lambda_C-\lambda_A} + R_A - R_B \quad (20)$$

where, RED:665 nm, RE2:740 nm, RE3:783 nm, NIR: 833 nm, NNIR: 865 nm, and SWIR:1614 nm.

NDVI was developed in radiance data that were corrected only for the sun angle (Rouse *et al.* 1973) and HI can be also applied on radiance data (Kühn *et al.* 2004). In addition, FAI is considered less sensitive than NDVI to atmospheric conditions or solar/viewing geometry (Hu *et al.* 2015). However, FDI and the S2-based index were developed using atmospherically corrected data by the ACOLITE algorithm. In this study, atmospheric correction was not performed on the fused images before applying the indexes.

Figures 19(a-e)-20(a-e) present the implemented indexes. For the marine plastic litter detection, thresholds were roughly defined according to the range of the target values in order to delineate a significant percentage of the targets. The accuracy score was calculated for every index (Table 5). It is noted that

the recall and precision scores were not shown as they would be misleading since the values are similar for all indexes. This is expected because of the small number of target pixels (low precision) and our thresholding strategy (high recall).

Table 5
Accuracy score of plastic target detection on indexes images

Date	Method	TP	FP	FN	TN	Accuracy
2018	FDI	18	431	0	814130	0.999
	S2_based	12	13390	6	801171	0.984
	HI 789	18	2623	0	811938	0.997
	NDVI	18	3783	0	810778	0.995
	FAI	18	2952	0	811609	0.996
2021	FDI	191	1179	31	563481	0.998
	S2_based	191	21483	31	543177	0.962
	HI 789	217	517170	5	47490	0.084
	NDVI	222	148627	0	416033	0.737
	FAI	215	226014	7	338646	0.600

All indexes showed high accuracy score values in the first experiment with FDI producing the lowest number of false positive pixels. In the second experiment FDI showed highest accuracy and smallest number of false positive pixels. The lower performance of S2-based, HI, NDVI, and FAI in the second experiment could be explained by the low pixel coverage of the small plastic targets (1/3 of the 2 m fused pixel) and varying experimental conditions.

Figures 19(f)-20(f) show the results after implementing the FDI thresholds. It is observed that all plastic targets are detected, except for one small (0.6 × 0.6 m^2) target (PET), which was not clearly discriminated both in the fused and the original WV image. False detection of very shallow water (right next to the shoreline) and other floating objects (e.g., boats) are also observed. Finally, some false positives can be seen over the border of the wooden target. The rest of the indexes also falsely detected wood pixels (FDI: 60 px, S2-based:18 px, HI: 16 px, NDVI: 196 px, FAI: 63 px).

It should be noted that although FDI and the S2-based index were developed using data from the first experiment, thus a positive bias is expected, it is important to take into account the high performance in the second experiment where the properties of the targets and the environment differed. In contrast to the first experiment, the second employed a) PS, colourful HDPE and mixed plastic targets, b) greater variety of sizes, c) a wooden target, and d) different location (gulf vs bay). In addition, the positive bias is reduced by the fact that atmospheric corrections were not applied in our study contrary to the development of both indexes.

4. Conclusion

In this study, we fused Sentinel-2 (S2) and WorldView-2/3 (WV-2/3) images to increase the capability of S2 imagery to detect marine plastic targets. Although the plastics might be visible on the very high resolution (VHR) images, the SWIR region of the spectrum, which is captured in S2 imagery, is very important in order for the plastics to be identified and

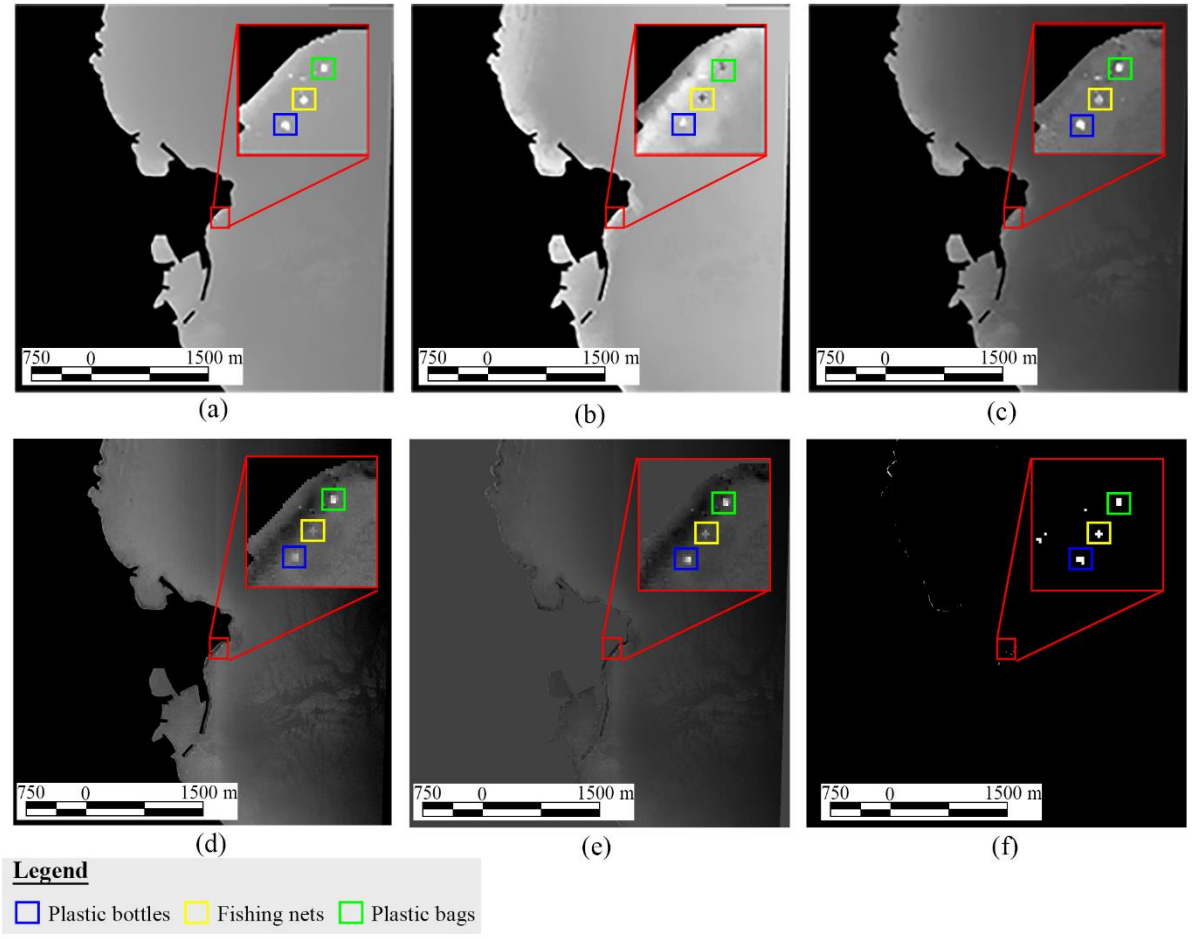

Fig. 19. Application of the marine litter indexes on the S2-WV-3 fused image collected on 07/06/2018 (land is removed) with the area of the plastic targets highlighted. (a) FDI. (b) S2-based. (c) HI. (d) NDVI. (e) FAI. (f) FDI after thresholding implementation.

distinguished from other materials. Artificial plastic targets were constructed and various image fusion techniques were evaluated for their performance in terms of preserving spectral and spatial information. The CNMF method proved the best for this application as it produces a fused image with clear edges, no blurring, and with a relative favourable impact on the spectral characteristics of the materials. DL methods showed high performance in terms of spectral similarity between the fused and the S2 images. In this regard, Fusion-GAN and Fusion-ResNet outperformed all other fusion methods (non-DL and DL). Among the plastic targets, only the small PET target was not distinguishable by visual inspection. It is worth noting that this target was not distinguishable even in the WV-2 image. This could be rather attributed to the transparency of the bottles that served as PET material for this target in combination with its placement in quite shallow seawater. The smallest plastic target that was observed in the fused image is 0.6×0.6 m^2 in size and is equivalent to 3% pixel coverage of the original S2 (20m) image and almost 1/3 of the pixel coverage of the original WV-2 (2m) image.

Furthermore, results showed that the VNIR (Visible-Near Infrared) combination is the most efficient for the image fusion. This is an important finding, because most satellite sensors are sensitive to the VNIR part of the spectrum, so the likelihood of temporally close acquisitions of the same marine litter accumulation increases, which is a critical factor considering that marine litter accumulations move, change or vanish in relatively short time. However, exactly for this reason, the described fusion approach might be limited for detecting litter accumulations in the sea, unless suitable constellation-based solutions can offer the possibility of "nearly" simultaneous image acquisitions between different satellites. It is worth mentioning that, still remaining in the frame of monitoring plastic pollution in the environment, the proposed fusion approach could find application in the detection of accumulations of plastic litter on land close to water bodies (i.e. sources of marine litter), which is a more static scenario, although with higher complexity in terms of background.

Five indexes that detect plastic material using discriminative features in the VIS, NIR, and SWIR parts of the spectrum were applied on the fused images. In this case, the SWIR bands of the fused image proved to be quite useful. FDI which exploits SWIR bands, was the only index that adequately detected plastic material in the fused images. It outperformed the other indexes, not only in the first experiment where indeed there was a positive bias but also in the second experiment. This finding reinforces the value of S2 SWIR bands for marine litter detection and shows that fusion with VHR images can produce fused S2 products able to detect marine plastic litter accumulations at smaller scales.

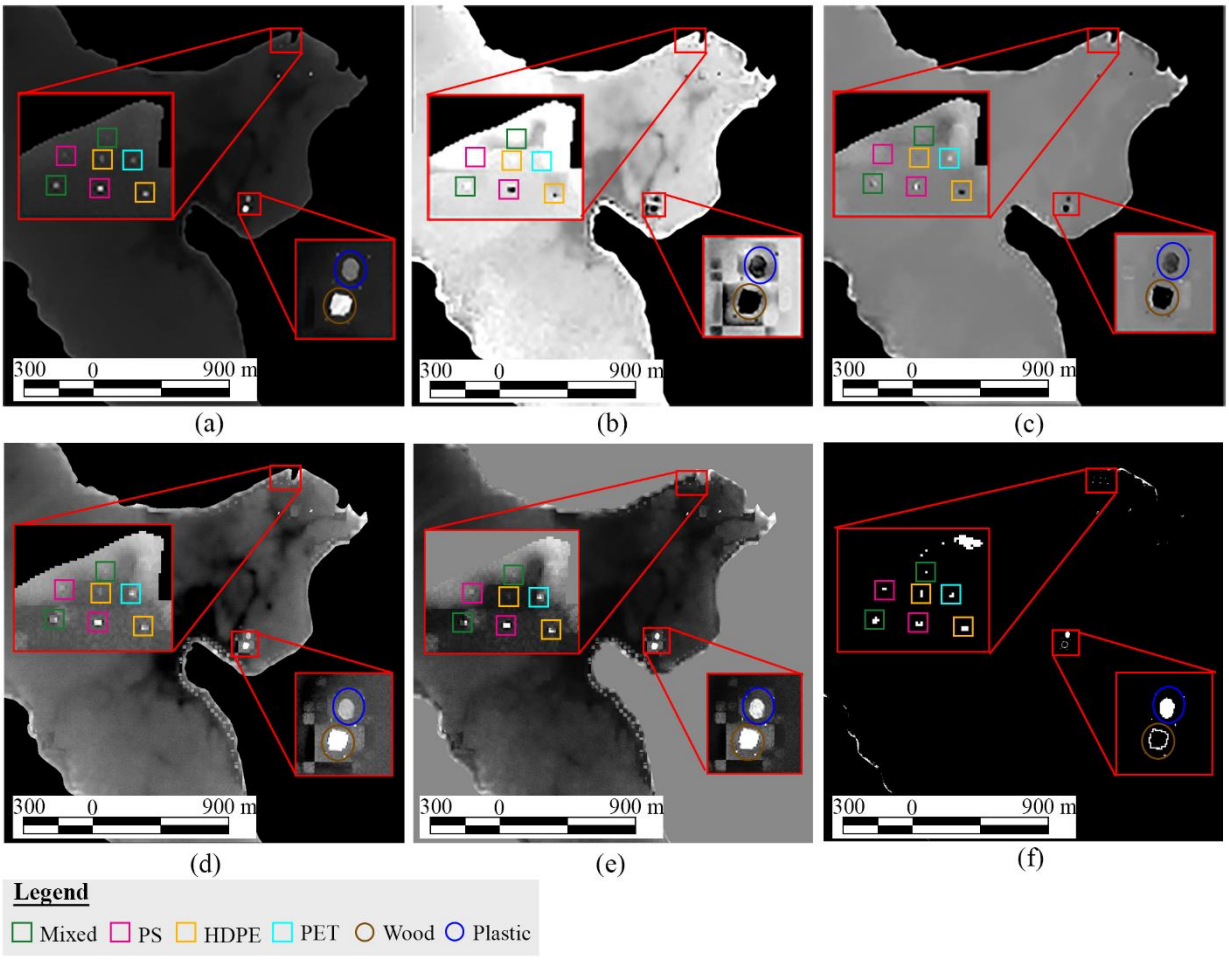

Fig. 20. Application of the marine litter indexes on the S2-WV-3 fused image collected on 21/06/2021 (land is removed) with the area of the plastic targets highlighted. (a) FDI. (b) S2-based. (c) HI. (d) NDVI. (e) FAI. (f) FDI after thresholding implementation.

Concerning the computational needs, they would not prevent the usage of our approach in an operational level because current mainstream industrial and research hardware is capable of maintaining the computational needs for the fusion methods implemented in the study, even for the DL methods. Regarding costs, even though the required processing equipment can be considered inexpensive, procurement of the required commercial VHR data indeed increases the cost, at least for the moment.

Another interesting finding of this research was the observation of dissimilarities in the spectral regions of S2 bands between the signatures of the various plastic materials and other floating debris extracted from the fused images. This topic along with pinpointing the scalability of the proposed methodology, will be the subject of future work.

Acknowledgements

This work was implemented in the framework of the project REACT. The project is funded by the Discovery Element of the European Space Agency's Basic Activities (ESA contract no. 4000131235/20/NL/GLC). Also, it was partially supported by the Discovery Element of the European Space Agency's Basic Activities, Project: "Plastic Litter Project - Detection and monitoring of artificial plastic targets with satellite imagery and UAV" (ESA contract no. 4000131040/20/NL/GLC).

References

Acuña-Ruz, T., Uribe, D., Taylor, R., Amézquita, L., Guzmán, M.C., Merrill, J., Martínez, P., Voisin, L., Mattar B., C., 2018. Anthropogenic marine debris over beaches: Spectral characterization for remote sensing applications. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2018.08.008

Afonso, M. V., Bioucas-Dias, J.M., Figueiredo, M.A.T., 2011. An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems. IEEE Trans. Image Process. https://doi.org/10.1109/TIP.2010.2076294

Agarap, A.F., 2018. Deep Learning using Rectified Linear Units (ReLU) 2–8.

Akhtar, N., Shafait, F., Mian, A., 2015. Bayesian sparse representation for hyperspectral image super resolution, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2015.7298986

Basu, B., Sannigrahi, S., Basu, A.S., Pilla, F., 2021. Development of novel classification algorithms for detection of floating plastic debris in coastal waterbodies using multispectral sentinel-2 remote sensing imagery. Remote Sens. https://doi.org/10.3390/rs13081598

Berné, O., Helens, A., Pilleri, P., Joblin, C., 2010. Non-negative matrix factorization pansharpening of hyperspectral data: An application to mid-infrared astronomy, in: 2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing. pp. 1–4. https://doi.org/10.1109/WHISPERS.2010.5594900

Bieniarz, J., Cerra, D., Avbelj, J., Reinartz, P., 2011. Hyperspectral Image Resolution Enhancement Based on Spectral Unmixing and Information Fusion. ISPRS Hann. Work. 2011 High-Resolution Earth Imaging Geospatial Inf. 1, 1–6. https://doi.org/10.5194/isprsarchives-XXXVIII-4-W19-33-2011

Biermann, L., Clewley, D., Martinez-Vicente, V., Topouzelis, K., 2020. Finding Plastic Patches in Coastal Waters using Optical Satellite Data. Sci. Rep. https://doi.org/10.1038/s41598-020-62298-z

Bioucas-Dias, J.M., 2009. A variable splitting augmented Lagrangian approach to linear spectral unmixing, in: 2009 First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing. IEEE, pp. 1–4. https://doi.org/10.1109/WHISPERS.2009.5289072

Casabianca, S., Capellacci, S., Giacobbe, M.G., Dell'Aversano, C., Tartaglione, L., Varriale, F., Narizzano, R., Risso, F., Moretto, P., Dagnino, A., Bertolotto, R., Barbone, E., Ungaro, N., Penna, A., 2019. Plastic-associated harmful microalgal assemblages in marine environment. Environ. Pollut. https://doi.org/10.1016/j.envpol.2018.09.110

Chamas, A., Moon, H., Zheng, J., Qiu, Y., Tabassum, T., Jang, J.H., Abu-Omar, M., Scott, S.L., Suh, S., 2020. Degradation Rates of Plastics in the Environment. ACS Sustain. Chem. Eng. https://doi.org/10.1021/acssuschemeng.9b06635

Fei-Fei, L., Deng, J., Li, K., 2010. ImageNet: Constructing a large-scale image database. J. Vis. 9, 1037–1037. https://doi.org/10.1167/9.8.1037

Garaba, S.P., Dierssen, H.M., 2018. An airborne remote sensing case study of synthetic hydrocarbon detection using short wave infrared absorption features identified from marine-harvested macro- and microplastics. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2017.11.023

Gargiulo, M., Mazza, A., Gaetano, R., Ruello, G., Scarpa, G., 2019. Fast super-resolution of 20 m Sentinel-2 bands using convolutional neural networks. Remote Sens. 11, 1–18. https://doi.org/10.3390/rs11222635

GESAMP, 2019. Guidelines for the monitoring and assessment of plastic litter in the ocean, Reports and Studies GESAMP.

Goddijn-Murphy, L., Dufaur, J., 2018. Proof of concept for a model of light reflectance of plastics floating on natural waters. Mar. Pollut. Bull. 135, 1145–1157. https://doi.org/10.1016/j.marpolbul.2018.08.044

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative Adversarial Networks. Commun. ACM 63, 139–144. https://doi.org/10.1145/3422622

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-Decem, 770–778. https://doi.org/10.1109/CVPR.2016.90

Hu, C., 2009. A novel ocean color index to detect floating algae in the global oceans. Remote Sens. Environ. 113, 2118–2129. https://doi.org/10.1016/j.rse.2009.05.012

Hu, C., Feng, L., Hardy, R.F., Hochberg, E.J., 2015. Spectral and spatial requirements of remote measurements of pelagic Sargassum macroalgae. Remote Sens. Environ. 167, 229–246. https://doi.org/10.1016/j.rse.2015.05.022

Hu, C., Lee, Z., Franz, B., 2012. Chlorophyll a algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. J. Geophys. Res. Ocean. 117. https://doi.org/10.1029/2011JC007395

Jagalingam, P., Hegde, A.V., 2015. A Review of Quality Metrics for Fused Image. Aquat. Procedia 4, 133–142. https://doi.org/https://doi.org/10.1016/j.aqpro.2015.02.019

Jambeck, J.R., Geyer, R., Wilcox, C., Siegler, T.R., Perryman, M., Andrady, A., Narayan, R., Law, K.L., 2015. Plastic waste inputs from land into the ocean. Science (80-.). 347, 768–771. https://doi.org/10.1126/science.1260352

Kikaki, A., Karantzalos, K., Power, C.A., Raitsos, D.E., 2020. Remotely sensing the source and transport of marine plastic debris in Bay Islands of Honduras (Caribbean Sea). Remote Sens. 12. https://doi.org/10.3390/rs12111727

Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization 1–15.

Kotwal, K., Chaudhuri, S., 2013. A novel approach to quantitative evaluation of hyperspectral image fusion techniques. Inf. Fusion. https://doi.org/10.1016/j.inffus.2011.03.008

Kremezi, M., Karathanassi, V., 2019. Correcting the BRDF effects on Sentinel-2 ocean images, in: RSCy2019 - Seventh International Conference on Remote Sensing and Geoinformation of the Environment. https://doi.org/10.1117/12.2533653

Kremezi, M., Kristollari, V., Karathanassi, V., Topouzelis, K., Kolokoussis, P., Taggio, N., Aiello, A., Ceriola, G., Barbone, E., Corradi, P., 2021. Pansharpening PRISMA Data for Marine Plastic Litter Detection Using Plastic Indexes. IEEE Access. https://doi.org/10.1109/ACCESS.2021.3073903

Kuester, M., 2016. Radiometric Use of WorldView-3 Imagery. Tech. Note.

Kühn, F., Oppermann, K., Hörig, B., 2004. Hydrocarbon Index – an algorithm for hyperspectral detection of hydrocarbons. Int. J. Remote Sens. 25, 2467–2473. https://doi.org/10.1080/01431160310001642287

Lanaras, C., Baltsavias, E., Schindler, K., 2017. Hyperspectral super-resolution with spectral unmixing constraints. Remote Sens. https://doi.org/10.3390/rs9111196

Lanaras, C., Baltsavias, E., Schindler, K., 2015. Hyperspectral super-resolution by coupled spectral unmixing, in: Proceedings of the IEEE International Conference on Computer Vision. https://doi.org/10.1109/ICCV.2015.409

Lanaras, C., Bioucas-Dias, J., Galliani, S., Baltsavias, E., Schindler, K., 2018. Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. ISPRS J. Photogramm. Remote Sens. 146, 305–319. https://doi.org/10.1016/j.isprsjprs.2018.09.018

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., Shi, W., 2017. Photo-realistic single image super-resolution using a generative adversarial network. Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 2017-Janua, 105–114. https://doi.org/10.1109/CVPR.2017.19

Lee, D.D., Seung, H.S., 1999. Learning the parts of objects by non-negative matrix factorization. Nature 401, 788–791. https://doi.org/10.1038/44565

Li, Y., Li, B., 2021. Super-Resolution of sentinel-2 images at 10m resolution without reference images 1–22. https://doi.org/10.20944/preprints202104.0556.v1

Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M., 2017. Enhanced Deep Residual Networks for Single Image Super-Resolution. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work. 2017-July, 1132–1140. https://doi.org/10.1109/CVPRW.2017.151

Loncan, L., De Almeida, L.B., Bioucas-Dias, J.M., Briottet, X., Chanussot, J., Dobigeon, N., Fabre, S., Liao, W., Licciardi, G.A., Simoes, M., Tourneret, J.Y., Veganzones, M.A., Vivone, G., Wei, Q., Yokoya, N., 2015. Hyperspectral Pansharpening: A Review. IEEE Geosci. Remote Sens. Mag. https://doi.org/10.1109/MGRS.2015.2440094

Martinez-Vicente, V., Biermann, L., Mata, A., 2020. Optical Methods for Marine Litter Detection (OPTIMAL) - Final Report. https://doi.org/https://doi.org/10.5281/ZENODO.3748797

Masi, G., Cozzolino, D., Verdoliva, L., Scarpa, G., 2016. Pansharpening by convolutional neural networks. Remote Sens. https://doi.org/10.3390/rs8070594

Maximenko, N., Corradi, P., Law, K.L., Sebille, E. Van, Garaba, S.P., Lampitt, R.S., Galgani, F., Martinez-Vicente, V., Goddijn-Murphy, L., Veiga, J.M., Thompson, R.C., Maes, C., Moller, D., Löscher, C.R., Addamo, A.M., Lamson, M., Centurioni, L.R., Posth, N., Lumpkin, R., Vinci, M., Martins, A.M., Pieper, C.D., Isobe, A., Hanke, G., Edwards, M., Chubarenko, I.P., Rodriguez, E., Aliani, S., Arias, M., Asner, G.P., Brosich, A., Carlton, J.T., Chao, Y., Cook, A.M., Cundy, A., Galloway, T.S., Giorgetti, A., Goni, G.J., Guichoux, Y., Hardesty, B.D., Holdsworth, N., Lebreton, L., Leslie, H.A., Macadam-Somer, I., Mace, T., Manuel, M., Marsh, R., Martinez, E., Mayor, D., Le Moigne, M., Jack, M.E.M., Mowlem, M.C., Obbard, R.W., Pabortsava, K., Robberson, B., Rotaru, A.E., Spedicato, M.T., Thiel, M., Turra, A., Wilcox, C., 2019. 'Towards the integrated marine debris observing system'. Front. Mar. Sci. 6. https://doi.org/10.3389/fmars.2019.00447

Nascimento, J.M.P., Dias, J.M.B., 2005. Vertex component analysis: a fast algorithm to unmix hyperspectral data. IEEE Trans. Geosci. Remote Sens. 43, 898–910. https://doi.org/10.1109/TGRS.2005.844293

Nwankpa, C., Ijomah, W., Gachagan, A., Marshall, S., 2018. Activation Functions: Comparison of trends in Practice and Research for Deep Learning 1–20.

Palsson, F., Sveinsson, J.R., Ulfarsson, M.O., 2018. Sentinel-2 image fusion using a deep residual network. Remote Sens. 10. https://doi.org/10.3390/rs10081290

Papageorgiou, D., 2019. 'Floating plastic detection using Sentinel-2 imagery. 'Diploma thesis, Supervised by V. Karathanassi, National Technical University of Athens, Athens, Greece.

PLP, P.L.P., 2021. Plastic Litter Project 2021 [WWW Document].

Rios, L.M., Moore, C., Jones, P.R., 2007. Persistent organic pollutants carried by synthetic polymers in the ocean environment. Mar. Pollut. Bull. 54, 1230–1237. https://doi.org/10.1016/j.marpolbul.2007.03.022

Romero, L.S., Marcello, J., Vilaplana, V., 2020. Super-resolution of Sentinel-2 imagery using generative adversarial networks. Remote Sens. 12, 1–25. https://doi.org/10.3390/RS12152424

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-319-24574-4_28

Rouse, J.W. , Haas, R.H., SChell, J.A., Deering, D.W., 1973. Monitoring vegetation systems in the great plains with ERTS, in: Third Earth Resources Technology Satellite-1 Symposium. pp. 309–317.

Salgado-Hernanz, P.M., Bauzà, J., Alomar, C., Compa, M., Romero, L., Deudero, S., 2021. Assessment of marine litter through remote sensing: recent approaches and future goals. Mar. Pollut. Bull. https://doi.org/10.1016/j.marpolbul.2021.112347

Salgueiro, L., Marcello, J., Vilaplana, V., 2021. Single-image super-resolution of sentinel-2 low resolution bands with residual dense convolutional neural networks. Remote Sens. 13, 1–20. https://doi.org/10.3390/rs13245007

Scarpa, G., Member, Senior, Vitale, S., Member, Student, Cozzolino, D., 2018. Target-Adaptive CNN-Based Pansharpening 56, 5443–5457.

Simoes, M., Bioucas-Dias, J., Almeida, L.B., Chanussot, J., 2015. A convex formulation for hyperspectral image superresolution via subspace-based regularization. IEEE Trans. Geosci. Remote Sens. https://doi.org/10.1109/TGRS.2014.2375320

Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc. 1–14.

Smail, E.A., DiGiacomo, P.M., Seeyave, S., Djavidnia, S., Celliers, L., Le Traon, P.Y., Gault, J., Escobar-Briones, E., Plag, H.P., Pequignet, C., Bajona, L., Zhang, L., Pearlman, J., Steven, A., Hodge, J., Racault, M.F., Storlazzi, C., Skirving, W., Hoeke, R., Marra, J., van Dongeren, A., Muller-Karger, F., Cripe, D., Takaki, D., 2019. 'An introduction to the 'Oceans and Society: Blue Planet initiative'. J. Oper. Oceanogr. https://doi.org/10.1080/1755876X.2019.1634959

Topouzelis, K., Papageorgiou, D., Karagaitanakis, A., Papakonstantinou, A., Ballesteros, M.A., 2020. Remote sensing of sea surface artificial floating plastic targets with Sentinel-2 and unmanned aerial systems (plastic litter project 2019). Remote Sens. https://doi.org/10.3390/rs12122013

Topouzelis, K., Papageorgiou, D., Suaria, G., Aliani, S., 2021. Floating marine litter detection algorithms and techniques using optical remote sensing data: A review. Mar. Pollut. Bull. https://doi.org/10.1016/j.marpolbul.2021.112675

Topouzelis, K., Papakonstantinou, A., Garaba, S.P., 2019. Detection of floating plastics from satellite and unmanned aerial systems (Plastic Litter Project 2018). Int. J. Appl. Earth Obs. Geoinf. 79, 175–183. https://doi.org/10.1016/j.jag.2019.03.011

UNEP, United Nations Environment Programme., 2016. 'Marine plastic debris and microplastics – Global lessons and research to inspire action and guide policy change.', Nairobi.

UNEP, 2005. ' Marine litter, an analytical overview, The right start to a healthy life : levelling-up the health gradient among children, young people and families in the European Union : what works?'. 123–152.

Updike, T., Comp, C., 2010. Radiometric Use of WorldView-2 Imagery Technical Note. DigitalGlobe.

Wang, M., Hu, C., 2016. Mapping and quantifying Sargassum distribution and coverage in the Central West Atlantic using MODIS observations. Remote Sens. Environ. 183, 350–367. https://doi.org/10.1016/j.rse.2016.04.019

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Loy, C.C., 2019. ESRGAN: Enhanced super-resolution generative adversarial networks. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 11133 LNCS, 63–79. https://doi.org/10.1007/978-3-030-11021-5_5

Wei, Q., Bioucas-Dias, J., Dobigeon, N., Tourneret, J.Y., Chen, M., Godsill, S., 2016. Multiband Image Fusion Based on Spectral Unmixing, in: IEEE Transactions on Geoscience and Remote Sensing. https://doi.org/10.1109/TGRS.2016.2598784

Wei, Q., Dobigeon, N., Tourneret, J.Y., 2015a. Bayesian Fusion of Multi-Band Images. IEEE J. Sel. Top. Signal Process. https://doi.org/10.1109/JSTSP.2015.2407855

Wei, Q., Dobigeon, N., Tourneret, J.Y., 2015b. Fast Fusion of Multi-Band Images Based on Solving a Sylvester Equation. IEEE Trans. Image Process. https://doi.org/10.1109/TIP.2015.2458572

Wu, J., He, Z., Hu, J., 2020. Sentinel-2 sharpening via parallel residual network. Remote Sens. 12. https://doi.org/10.3390/rs12020279

Yang, J., Fu, X., Hu, Y., Huang, Y., Ding, X., Paisley, J., 2017. PanNet: A Deep Network Architecture for Pan-Sharpening. Proc. IEEE Int. Conf. Comput. Vis. 2017-Octob, 1753–1761. https://doi.org/10.1109/ICCV.2017.193

Yokoya, N., Grohnfeldt, C., Chanussot, J., 2017. Hyperspectral and multispectral data fusion: A comparative review of the recent literature. IEEE Geosci. Remote Sens. Mag. https://doi.org/10.1109/MGRS.2016.2637824

Yokoya, N., Yairi, T., Iwasaki, A., 2012. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. IEEE Trans. Geosci. Remote Sens. https://doi.org/10.1109/TGRS.2011.2161320

Yokoya, N., Yairi, T., Iwasaki, A., 2011. Coupled non-negative matrix factorization (CNMF) for hyperspectral and multispectral data fusion: Application to pasture classification, in: International Geoscience and Remote Sensing Symposium (IGARSS). https://doi.org/10.1109/IGARSS.2011.6049465

Zhang, K., Sumbul, G., Demir, B., 2020. An Approach to Super-Resolution of Sentinel-2 Images Based on Generative Adversarial Networks. 2020 Mediterr. Middle-East Geosci. Remote Sens. Symp. M2GARSS 2020 - Proc. 69–72. https://doi.org/10.1109/M2GARSS47143.2020.9105165

Zhang, Y., Li, Kunpeng, Li, Kai, Wang, L., Zhong, B., Fu, Y., 2018. Image super-resolution using very deep residual channel attention networks. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 11211 LNCS, 294–310. https://doi.org/10.1007/978-3-030-01234-2_18

Zhu, X., Xu, Y., Wei, Z., 2019. Super-Resolution of Sentinel-2 Images Based on Deep Channel-Attention Residual Network, in: IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 628–631. https://doi.org/10.1109/IGARSS.2019.8897860