

DEEPSUN

Interconnection and valorisation of long-term solar datasets via deep learning

Contract - B2/191/P2/DeepSun

SAMENVATTING

CONTEXT

Zonnevlekken zijn donkere, ‘koude’ plekken die in groepen op het oppervlak van de zon verschijnen als een uiting van zonnemagnetisme. Het magnetische veld dat in zonnevlekken is ingebed, is de drijvende kracht achter de variabiliteit van de zon die het ruimteweer rondom de aarde dagelijks beïnvloedt. Het bestuderen van de evolutie van zonnevlekken op lange termijn is van groot belang voor verschillende aspecten van de zonnefysica, van helioseismologie tot het modelleren van de zonnestraling en de voorspelling van ruimteweerfenomenen.

De Koninklijke Sterrenwacht van België (KSB) speelt een belangrijke rol in waarnemingen van zonnevlekken: in 1939 startte het KSB-zonnewaarnemingsstation (genaamd 'USET', voor Uccle Solar Equatorial Table) een waarnemingsprogramma in samenwerking met de sterrenwacht van Zürich. Dit programma bestaat uit dagelijkse tekeningen van de zonnevlekkenconfiguratie; deze collectie wordt nog dagelijks aangevuld. Daarnaast produceert de USET-faciliteit sinds 2002 witlichtbeelden (WL) van de fotosfeer en sinds 2012 beelden van de chromosfeer in de Ca IJK-band. De co-temporele en co-ruimtelijke acquisitie van witlicht- en CaIJK-beelden maakt het mogelijk om deze datasets met elkaar te verbinden met behulp van nieuwe beeldverwerkingstechnieken.

DOEL EN METHODOLOGIE

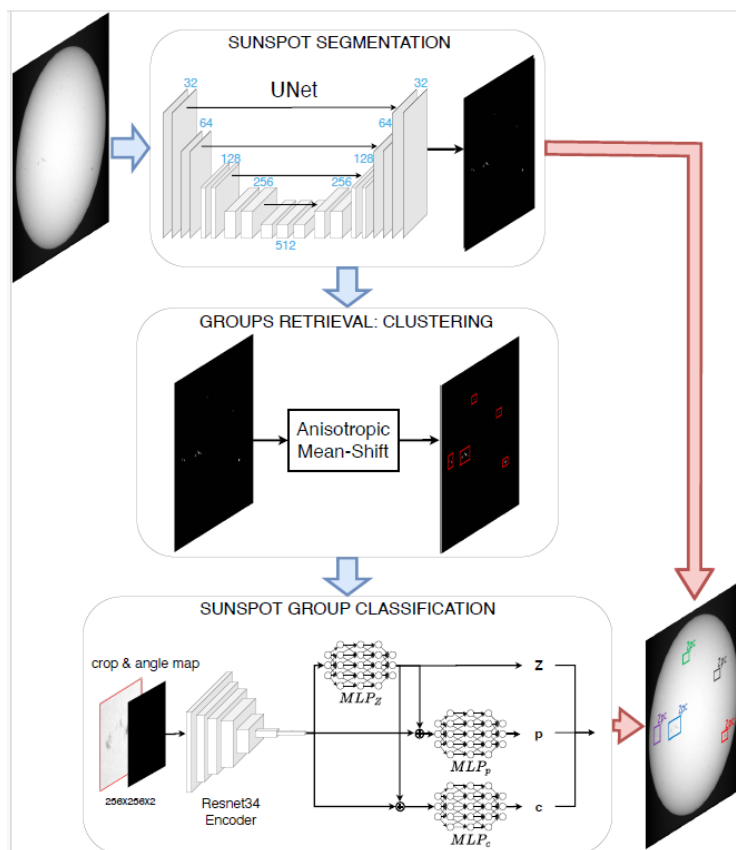
Het overkoepelende doel van dit project is het produceren van nieuwe producten voor de wetenschappelijke data-exploitatie en -verspreiding vanuit de verschillende USET-datasets.

Ons eerste doel is het automatisch detecteren, groeperen en classificeren van zonnevlekkengroepen op basis van enkel de witlichtbeelden van USET. Een dergelijke classificatie biedt ondersteuning aan ruimteweervoorspellers en maakt systematische studies van zonnevlekken mogelijk.

Om dit doel te bereiken, hebben we de SunSCC-pijplijn ontwikkeld, zoals weergegeven in Figuur 1. Deze pijplijn neemt USET-witlichtbeelden als input en levert als output segmentatiemaskers van individuele zonnevlekken, samen met een omkadring van de overeenkomstige groepen en hun McIntosh-classificatie, voorzien van betrouwbaarheidsscores. De pijplijn is onderverdeeld in drie onafhankelijke blokken:

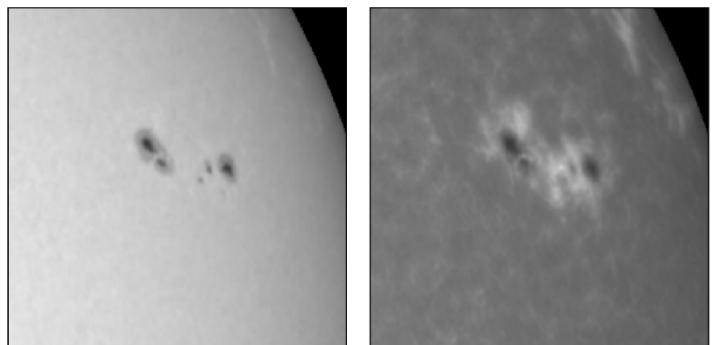
1. Een segmentatieblok voor zonnevlekken, gebaseerd op een U-net-architectuur opgebouwd uit convolutionele neurale netwerken (CNN), die zich doeltreffend heeft bewezen bij segmentatietaken.
2. Een aggregatieblok dat individuele zonnevlekken groepeerd tot een zonnevlekkengroep, op basis van een aangepast mean-shift-algoritme.
3. Een classificatieblok, opgebouwd rond multi-layer perceptrons (MLP), ontworpen om de McIntosh-classificatie — bestaande uit drie componenten — van de zonnevlekkengroep te voorspellen.

Ons tweede doel is het opvullen van hiaten in historische gegevens door gebruik te maken van de fysieke correlatie tussen verschillende observatiemodaliteiten. Vertrekkend van langlopende witlichtwaarnemingen van zonnevlekken, is ons doel om de overeenkomstige plages in Ca II-K-beelden te reconstrueren (zie Figuur 2). Om dit te bereiken, hebben we eerst oplossingen onderzocht op basis van generative adversarial networks (GANs), vanwege hun vermogen tot beeld-naar-beeldvertaling en dus, in ons geval, tot het genereren van realistische Ca II-K-beelden uit witlichtbeelden. We stelden echter vast dat deze methode fysiek niet-onderbouwde artefacten introduceert. Dit leidde ons ertoe om I2IwFILM voor te stellen, een nieuwe modelarchitectuur voor data-exploitatie in de zonnefysica. Door geavanceerde feature-extractie te combineren met feature map-modulatie, maakt deze methode een verrijkte reconstructie mogelijk van de doelmodaliteit (in dit geval Ca II-K) op basis van expliciet geëxtraheerde bronkenmerken.



Figuur 1. SunSCC-pijplijn voor segmentatie, clustering en classificatie van zonnevlekken. Beelden van de volledige zonneshijf met een resolutie van 2048×2048 worden onderverdeeld in een 4×4 -rooster, waarbij elk segment afmetingen heeft van 512×512 pixels. Een U-net-segmentatienetwerk voorspelt maskers voor elk segment, die vervolgens worden samengevoegd tot een segmentatiemasker van 2048×2048 . De gedetecteerde zonnevlekken worden gegroepeerd in zonnevlekkengroepen met behulp van een aangepast mean-shift-algoritme. Elke geïdentificeerde groep wordt, samen met een hoekafstandskaart, doorgegeven aan een classificatienetwerk bestaande uit een ResNet34-beeldencoder en drie MLP's, waarbij elk MLP gespecialiseerd is in de classificatie van één component van het McIntosh-systeem.

Figuur 2. Voorbeeld van een paar WL- en Ca II-K-subbeelden, die ruimtelijk zijn uitgelijnd en gecentreerd rond een actief gebied. Het WL-subbeeld toont zonnevlekken en gebieden rond de zonnevlekken die iets helderder zijn dan de achtergrond, de zogenaamde zonnefakkels. Deze zonnefakkels komen fysiek overeen met de heldere 'plages' die in het Ca II-K-subbeeld rond de donkere zonnevlekken worden waargenomen.



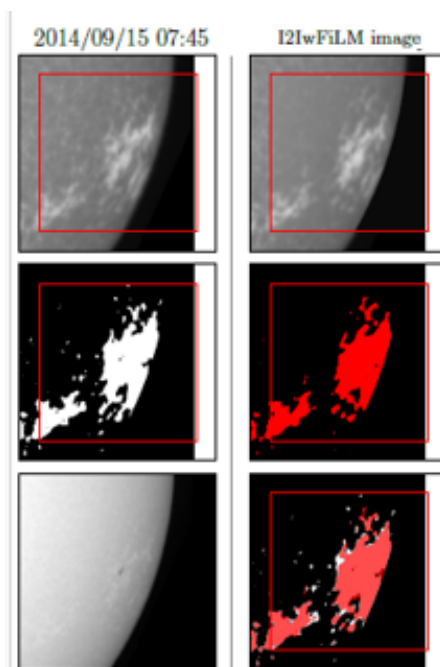
Ten slotte willen we dat onze resultaten gemakkelijk toegankelijk zijn voor de wetenschappelijke gemeenschap en dat burgers worden betrokken bij de verdere exploitatie van onze resultaten.

CONCLUSIES

Met betrekking tot ons eerste doel levert het segmentatiegedeelte van de SunSCC-pijplijn betrouwbare zonnevlekmaskers, zelfs wanneer atmosferische artefacten de USET-beelden degraderen. De ontwikkelde clusteringmethodologie maakt gebruik van de omvang van zonnevlekkengroepen en de grootte van individuele zonnevlekken om met succes de patronen uit de USET-catalogi te reproduceren. Ten slotte presteert de voorgestelde CNN-architectuur voor McIntosh-classificatie beter dan bestaande oplossingen die gebaseerd zijn op satellietbeelden, ondanks het feit dat hier uitsluitend gebruik wordt gemaakt van grondgebaseerde waarnemingen.

Met betrekking tot ons tweede doel hebben evaluaties aangetoond dat het door ons voorgestelde I2IwFiLM-schema fysiek consistente voorspellingen genereert, met minder hallucinaties dan op adversarial-netwerken gebaseerde methodes. Een voorbeeld van dergelijke vertalingen is weergegeven in Figuur 3. Deze synthetische beelden kunnen worden vergeleken met Ca IIK-waarnemingen uit historische data-archieven en kunnen helpen om eventuele inconsistenties in deze archieven te corrigeren.

Ten slotte hebben we USET-dataproducten via standaard virtuele observatoria verspreid onder de wetenschappelijke gemeenschap. Om de resultaten van DeepSun verder te verspreiden en te valoriseren, hebben we ook een burgerwetenschapsproject voorgesteld waarbij de taak bestaat uit het classificeren van zonnevlekken, met het oog op het vergelijken van fouten die worden gemaakt door geautomatiseerde en handmatige classificatiemethoden.



Figuur 3 Ca II K-reconstructies van plages nabij de rand, gemaakt op basis van waarnemingen op 15 september 2014 om 07:15 uur, met behulp van ons voorgestelde I2IwFiLM-algoritme. De bovenste rij toont het originele CaIIK-beeld (links) en de reconstructies door I2IwFiLM (rechts). De middelste rij toont segmentatiemaskers van plages die zichtbaar zijn in het originele CaIIK-beeld (links) en de overeenkomstige reconstructies door I2IwFiLM (rechts). De onderste rij toont het input-witlichtbeeld (WL) (links) en overlays van het doelplage-masker met de door het model gegenereerde plagemaskers van I2IwFiLM (rechts), wat een directe vergelijking mogelijk maakt. Het rode vierkant komt overeen met het subbeeld dat als input voor de modellen is gebruikt; er wordt een iets groter beeld getoond om de consistentie met de omgeving aan te geven.

SLEUTELWOORDEN

Zonnevlekkensegmentatie, zonnevlekkengroepering, zonnevlekkenclassificatie, beeld-naar-beeldvertaling, deep learning, convolutionele neurale netwerken