Providing KappaMask-based cloud and cloud segmentation masks for every Sentinel-2 product over Europe

Tetiana Shtym
with contributions from the KappaZeta team

tetiana.shtym@kappazeta.ee

"Use cases from the EOSC community", 16th May 2023
Overview of use case
Overview of use case
Comparison to other cloudmasks on the KappaSet test set

- True Color Image
- Label
- KappaMask L1C
- Sen2Cor

- Fmask
- IdePix
- MAJA
- S2Cloudless

Legend:
- Clear
- Semi-transparent
- Cloud shadow
- Cloud
- Missing
Processing time

Time comparison (in minutes) performed on the single Sentinel-2 Level-1C product inference. KappaMaskv2 L1C with GPU and CPU, Sen2Cor, Fmask and S2Cloudless on generating a 10 m resolution classification map. IdePix classification map is at 20 m resolution. Sen2Cor's 20m classification mask was resampled to 10 m.

<table>
<thead>
<tr>
<th></th>
<th>KappaMaskv2</th>
<th>KappaMaskv2</th>
<th>Sen2Cor v2.10</th>
<th>Fmask</th>
<th>IdePix Without cloud shadow processing</th>
<th>S2Cloudless</th>
</tr>
</thead>
<tbody>
<tr>
<td>Running time</td>
<td>02:58</td>
<td>06:01</td>
<td>05:50</td>
<td>06:10</td>
<td>16:53</td>
<td>18:10</td>
</tr>
</tbody>
</table>

Test computer:
CPU - Intel Core i7-8700K, 64GB of RAM,
GPU - NVIDIA GeForce GTX 1070 with 8GB of VRAM,
Linux Ubuntu 18.04.5 LTS (Bionic Beaver)
Overview of use case

- KappaMask is an AI-based cloud and cloud shadow processor for Sentinel-2.
- KappaMask has outperformed other cloud masks (e.g. Fmask, Sen2Cor, IdePix, S2Cloudless and MAJA) on the diverse and challenging test set.

By generating KappaMask-based cloud and cloud shadow segmentation mask for every Sentinel-2 product available in the European region and hosting the masks on CreoDIAS with a free licence, we provide ample opportunities for testing KappaMask performance for all interested parties.

Read more here: https://kappazeta.ee/blog/make-the-globe-cloud-free-with-kappamaskv2
KappaMaskv2 overall flow

Sentinel-2 Level-1C or Level-2A product resampled to 10m resolution

Splitting into subtiles

N = 484

Predict on each subtile (predictor - DeepLabv3+)

Mosaic

Cloud mask at 10 m resolution
Computing KappaMasks at CreoDIAS

- Architecture is a simple Python-RQ based task-worker scheme
- Easy to distribute processing steps between servers
- Can avoid back-pressure
- Multiple parallel processes for some steps
- Workers run in Docker

NVIDIA RTX A6000 24C (24G VRAM) ~ 3 minutes per product
CreoDIAS user experience

- We received a dedicated project environment in OpenStack.
- Setting up virtual machines, IP address, volumes, etc is straightforward.
- Very comfortable to have a network pre-defined for accessing Sentinel data – virtual machines have all mounted by default.

"The only thing that took time in the initial setup was figuring out that default ubuntu user is "eouser" and not "ubuntu"
Results

- **30389** Sentinel-2 products processed starting from January
- ~ 52 GB
Conclusions

• Utilized an OpenStack project with pre-configured networks for seamless access to necessary input data.
• Benefited from flexible computing resources, allowing for efficient processing.
• Engaged in discussions with CESNET to store the final product using STAC metadata, ensuring accessibility to third parties.
Conclusions

- Utilized an OpenStack project with pre-configured networks for seamless access to necessary input data.
- Benefited from flexible computing resources, allowing for efficient processing.
- Engaged in discussions with CESNET to store the final product using STAC metadata, ensuring accessibility to third parties.

Contact

- Tetiana Shtym
- tetiana.shtym@kappazeta.ee
Conclusions

- Utilized an OpenStack project with pre-configured networks for seamless access to necessary input data.
- Benefited from flexible computing resources, allowing for efficient processing.
- Engaged in discussions with CESNET to store the final product using STAC metadata, ensuring accessibility to third parties.

Contact

- Tetiana Shtym
- tetiana.shtym@kappazeta.ee