REMOTE SENSING DATA ANALYTICS WITH THE UDOCKER CONTAINER TOOL USING MULTI-GPU DEEP LEARNING SYSTEMS

Gabriele Cavallaro[1], Valentin Kozlov[2], Markus Goetz[2], Morris Riedel[1]

[1]Juelich Supercomputing Centre, Forschungszentrum Juelich, Germany
[2]Steinbuch Centre for Computing, Karlsruhe Institute of Technology, Germany

- **Multi-GPU and -user systems can tackle intensive computational big data problems**
- **Containers are strategies for packing, deploying and running isolated application processes**
- **Simplify the application build and deployment process without performance penalties**
- **uDocker: allows executing Docker containers without the necessity of administrative privileges**
- **More transparent to deploy for less tech-savvy researchers**

JÜLICH Forschungszentrum | JÜLICH SUPERCOMPUTING CENTRE

KIT Karlsruher Institut für Technologie

1. CONTAINERS

- **Lightweight virtualization technology**
- **Encapsulate system environments into standard units of software**
- **Portable, easy to build, small footprint, and low runtime overhead**

(a) can be realized by installing nvidia-docker runtime,
(b) experimental feature,
(c) container MPI version has to match the HPC one,
(d) number of high severity issues in Singularity

	Docker	Singularity	Shifter	Charlie Cloud	uDocker
Privilege model	Root daemon	SUID/ UserNS	SUID	UserNS	chroot-like
Current production Linux distros support	–	+	+	–	+
No privileged or trusted daemon	–	+	+	+	+
Access to the host filesystem	+	+	+	+	+
Support for GPU	– (a)	+ (b)	–	–	+ (b)
Support for MPI	+	+	+	+	+ (c)
Pulling from Docker Hub	+	+	+	+	+
No system admin intervention required	–	–	–	–	+
No escalation of permissions	–	+ (d)	+	+	+
Works with all HPC scheduler	–	+	–	+	+

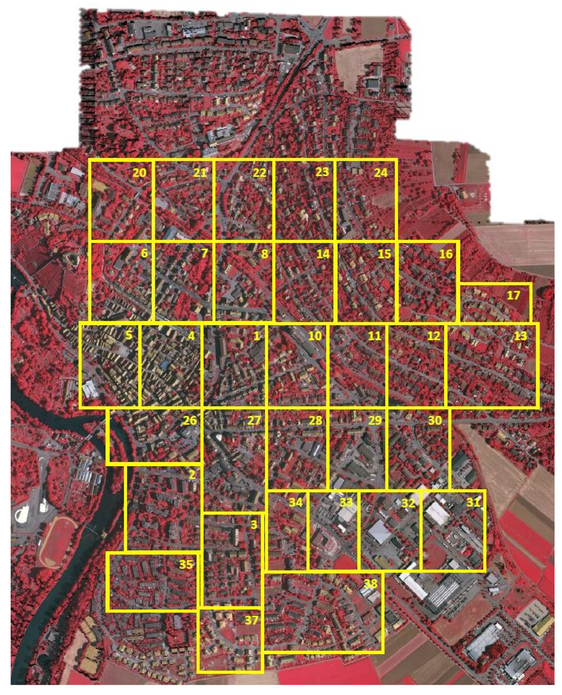
2. uDOCKER

UDOCKER

- **Docker requires the intervention of a privileged user, unacceptable in multi-user systems**
- **uDocker [1] allows running applications in a Docker container without root privileges and additional software**
- **Facilitates the deployment of new analytical models and workflows on multi-user systems**
- **Eases scientific reproducibility [2]**

3. EVALUATION

- **Res-Net-50 adapted into a fully convolutional network [3]**
- **2D semantic labeling with the Vaihingen dataset [4]**



4. RESULTS

- **No significant difference between *baremetal* and *uDocker***
- **Multiple GPUs speed up the training time**
- **Imperfect scaling due to Keras communication overhead**

CPU	RAM	NVIDIA GPU
2 x Intel Xeon E5 2630 v3	128 GB	4 x K80, 12 GB

	Training Time (s)	
Number of GPUs	baremetal	uDocker
1	3710 ± 10	3730 ± 10
2	2390 ± 30	2360 ± 16
4	1860 ± 40	1880 ± 10

References & Links:

[1] uDocker: https://github.com/indigo-dc/udocker
[2] Scientific Reproducibility: https://github.com/vykozlov/semseg-bids19 , https://hub.docker.com/r/vykozlov/semseg/
[3] Deep Network Model: https://www.azavea.com/blog/2017/05/30/deep-learning-on-aerial-imagery/
[4] Dataset: http://www2.isprs.org/commissions/comm3/wg4/2d-sem-label-vaihingen.html

Acknowledgments uDocker is being developed within DEEP HybridDataCloud project, which receives funding from the European Union's Horizon 2020 research and innovation programme under agreement RIA 777435. This project has received funding from the European Union's Horizon 2020 research and innovation program under the Grant Agreement No. 754304 DEEP-EST.

HELMHOLTZ ASSOCIATION