

# SD.Next on Linux

HW and OS level articles

- [SD.Next openvino docker compose](#)
- [SD.Next btrfs deduplication example](#)
- [Edit SD.Next task que using sqlite3](#)
- [SD.Next by liutyi docker compose](#)

# SD.Next openvino docker compose

## Install

```
vi Dockerfile
vi docker-compose.yml
docker compose up -d
docker logs sdnnext-openvino --follow
#docker compose down
```

## Dockerfile

### example without proxy on branch master

```
# SD.Next OpenVINO Dockerfile
# docs: <https://github.com/vladmandic/sdnnext/wiki/Docker>

# base image
FROM ubuntu:noble

# metadata
LABEL org.opencontainers.image.vendor="SD.Next"
LABEL org.opencontainers.image.authors="disty0"
LABEL org.opencontainers.image.url="https://github.com/vladmandic/sdnnext/"
LABEL
org.opencontainers.image.documentation="https://github.com/vladmandic/sdnnext/wiki/Docker"
LABEL org.opencontainers.image.source="https://github.com/vladmandic/sdnnext/"
LABEL org.opencontainers.image.licenses="AGPL-3.0"
LABEL org.opencontainers.image.title="SD.Next OpenVINO"
LABEL org.opencontainers.image.description="SD.Next: Advanced Implementation of Stable
Diffusion and other Diffusion-based generative image models"
LABEL org.opencontainers.image.base.name="https://hub.docker.com/_/ubuntu:noble"
LABEL org.opencontainers.image.version="latest"
```

```
# essentials
RUN apt-get update && \
    apt-get install -y --no-install-recommends --fix-missing \
    software-properties-common \
    build-essential \
    ca-certificates \
    wget \
    gpg \
    git

# intel compute runtime
RUN wget -q0 - https://repositories.intel.com/gpu/intel-graphics.key | gpg --yes --dearmor --
output /usr/share/keyrings/intel-graphics.gpg
RUN echo "deb [arch=amd64,i386 signed-by=/usr/share/keyrings/intel-graphics.gpg]
https://repositories.intel.com/gpu/ubuntu noble client" | tee /etc/apt/sources.list.d/intel-
gpu-noble.list
RUN apt-get update

RUN apt-get install -y --no-install-recommends --fix-missing \
    intel-opencl-icd \
    libze-intel-gpu1 \
    libze1

# required by pytorch / ipex
RUN apt-get install -y --no-install-recommends --fix-missing \
    libglib \
    libglib2.0-0 \
    libgomp1

# python3.12
RUN apt-get install -y --no-install-recommends --fix-missing \
    python3 \
    python3-dev \
    python3-venv \
    python3-pip

# cleanup
RUN /usr/sbin/ldconfig
RUN apt-get clean && rm -rf /var/lib/apt/lists/*
```

```

# stop pip and uv from caching
ENV PIP_NO_CACHE_DIR=true
ENV UV_NO_CACHE=true

# set paths to use with sdnex
ENV SD_DOCKER=true
ENV SD_DATADIR="/mnt/data"
ENV SD_MODELSDIR="/mnt/models"
ENV venv_dir="/mnt/python/venv"

# paths used by sdnex can be a volume if necessary
#VOLUME [ "/app" ]
#VOLUME [ "/mnt/data" ]
#VOLUME [ "/mnt/models" ]
#VOLUME [ "/mnt/python" ]
#VOLUME [ "/root/.cache/huggingface" ]

# git clone and run sdnex
RUN echo '#!/bin/bash\n git status || git clone https://github.com/vladmandic/sdnex.git
.\n/app/webui.sh "$@" | tee /bin/startup.sh
RUN chmod 755 /bin/startup.sh

# actually run sdnex
WORKDIR /app
ENTRYPOINT [ "startup.sh", "-f", "--use-openvino", "--uv", "--listen", "--insecure", "--
update", "--debug", "--api-log", "--log", "sdnex.log" ]

# expose port
EXPOSE 7860

# healthcheck function
# HEALTHCHECK --interval=60s --timeout=10s --start-period=60s --retries=3 CMD curl --fail
http://localhost:7860/sdapi/v1/status || exit 1

# stop signal
STOPSIGNAL SIGINT

```

## example with proxy

(to be removed or replaced to one is working) and on Developer branch

```
# SD.Next OpenVINO Dockerfile
# docs: <https://github.com/vladmandic/sdnext/wiki/Docker>

# base image
FROM ubuntu:noble

# metadata
LABEL org.opencontainers.image.vendor="SD.Next"
LABEL org.opencontainers.image.authors="disty0"
LABEL org.opencontainers.image.url="https://github.com/vladmandic/sdnext/"
LABEL
org.opencontainers.image.documentation="https://github.com/vladmandic/sdnext/wiki/Docker"
LABEL org.opencontainers.image.source="https://github.com/vladmandic/sdnext/"
LABEL org.opencontainers.image.licenses="AGPL-3.0"
LABEL org.opencontainers.image.title="SD.Next OpenVINO"
LABEL org.opencontainers.image.description="SD.Next: Advanced Implementation of Stable
Diffusion and other Diffusion-based generative image models"
LABEL org.opencontainers.image.base.name="https://hub.docker.com/_/ubuntu:noble"
LABEL org.opencontainers.image.version="latest"

# essentials
RUN apt-get update && \
    apt-get install -y --no-install-recommends --fix-missing \
    software-properties-common \
    build-essential \
    ca-certificates \
    wget \
    gpg \
    git

# intel compute runtime
RUN https_proxy=http://proxy.example.com:18080/ wget -q0 -
https://repositories.intel.com/gpu/intel-graphics.key | gpg --yes --dearmor --output
/usr/share/keyrings/intel-graphics.gpg
RUN echo "deb [arch=amd64,i386 signed-by=/usr/share/keyrings/intel-graphics.gpg]
https://repositories.intel.com/gpu/ubuntu noble client" | tee /etc/apt/sources.list.d/intel-
gpu-noble.list
RUN https_proxy=http://proxy.example.com:18080/ apt-get update
```

```
RUN https_proxy=http://proxy.example.com:18080/ apt-get install -y --no-install-recommends --
fix-missing \
    intel-opencl-icd \
    libze-intel-gpu1 \
    libze1

# required by pytorch / ipex
RUN https_proxy=http://proxy.example.com:18080/ apt-get install -y --no-install-recommends --
fix-missing \
    libglib \
    libglib2.0-0 \
    libgomp1

# python3.12
RUN https_proxy=http://proxy.example.com:18080/ apt-get install -y --no-install-recommends --
fix-missing \
    python3 \
    python3-dev \
    python3-venv \
    python3-pip

# cleanup
RUN /usr/sbin/ldconfig
RUN https_proxy=http://proxy.example.com:18080/ apt-get clean && rm -rf /var/lib/apt/lists/*

# stop pip and uv from caching
ENV PIP_NO_CACHE_DIR=true
ENV UV_NO_CACHE=true

# set paths to use with sdnex
ENV SD_DOCKER=true
ENV SD_DATADIR="/mnt/data"
ENV SD_MODELSDIR="/mnt/models"
ENV venv_dir="/mnt/python/venv"

# paths used by sdnex can be a volume if necessary
#VOLUME [ "/app" ]
#VOLUME [ "/mnt/data" ]
#VOLUME [ "/mnt/models" ]
```

```
#VOLUME [ "/mnt/python" ]
#VOLUME [ "/root/.cache/huggingface" ]

# git clone and run sdnex
RUN echo '#!/bin/bash\n git status || git clone https://github.com/vladmandic/sdnex.git -b dev
.\n/app/webui.sh "$@" | tee /bin/startup.sh
RUN chmod 755 /bin/startup.sh

# actually run sdnex
WORKDIR /app
ENTRYPOINT [ "startup.sh", "-f", "--use-openvino", "--uv", "--listen", "--insecure", "--
server-name sdnex.example.com", "--update", "--debug", "--api-log", "--log", "sdnex.log" ]

# expose port
EXPOSE 7860

# healthcheck function
# HEALTHCHECK --interval=60s --timeout=10s --start-period=60s --retries=3 CMD curl --fail
http://localhost:7860/sdapi/v1/status || exit 1

# stop signal
STOPSIGNAL SIGINT
```

## docker-compose.yml

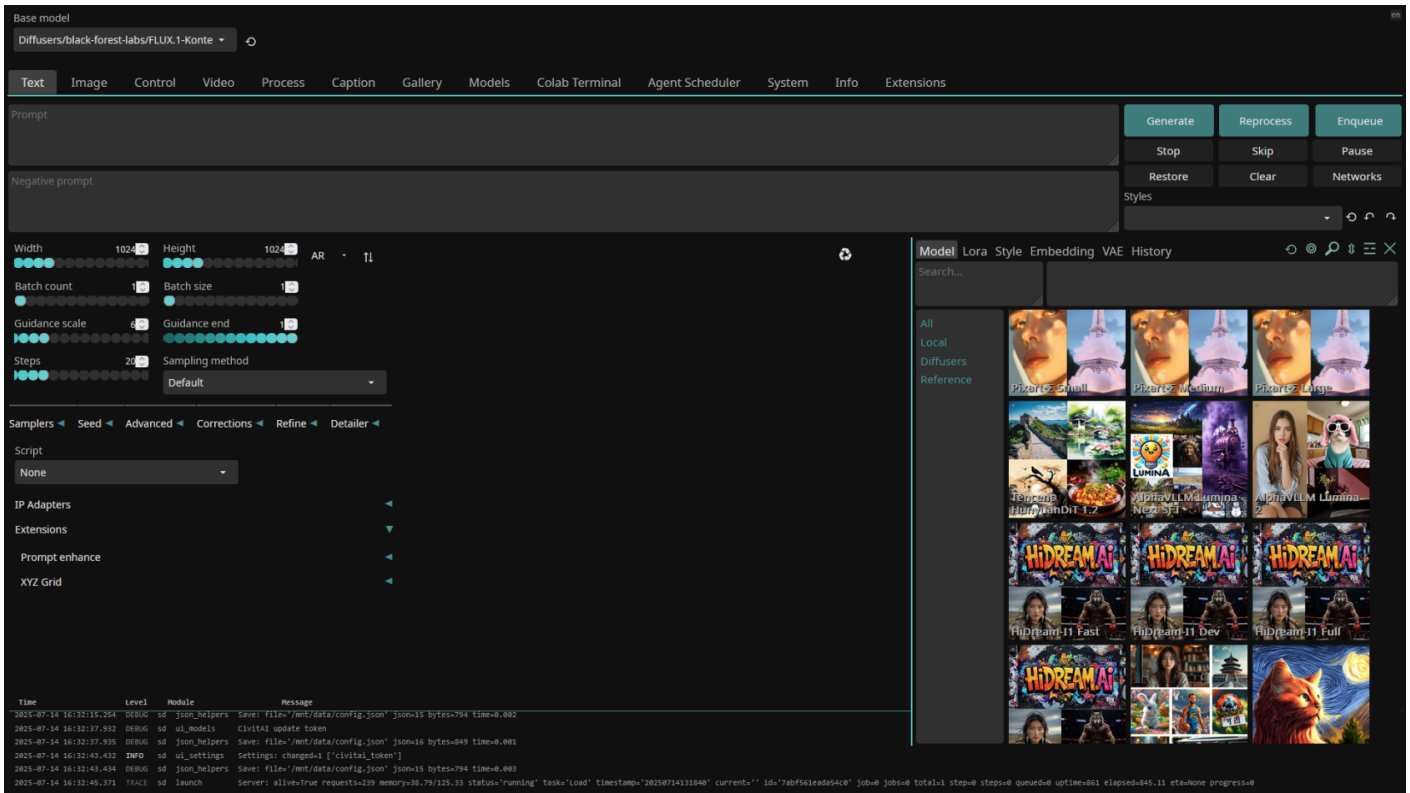
```
services:
  sdnex-openvino:
    build:
      dockerfile: Dockerfile
    image: sdnex-openvino:latest
    container_name: sdnex-openvino
    restart: unless-stopped
    devices:
      - /dev/dri:/dev/dri
# if Intel CPU got NPU
#   - /dev/accel:/dev/accel
# if need proxy for internet access
#   environment:
```

```
# - https_proxy=http://proxy.example.com:18080/  
ports:  
  - 7860:7860  
volumes:  
  - /docker/sdnext/app-volume:/app  
  - /docker/sdnext/mnt-volume:/mnt  
  - /docker/sdnext/huggingface-volume:/root/.cache/huggingface  
  - /dev/shm:/dev/shm
```

## Application screenshots



VERSION 79c0f8b1 2025-07-14 dev



# Post-install fixes

## add some python libraries

That is not something you need to do, but in rare case of it kind of missing in requirements.txt

```
# In case of some dependencies is missing (like sentencepiece in Dev branch at 2025-07-14)
# docker exec -it sdnnext-openvino bash
python3 -m venv /mnt/python/venv
/mnt/python/venv/bin/pip3 install sentencepiece
```

## Fix access to gated repo

(in case access is granted and token available)

## Access error to gated repo

```
sdnext-ipex | 20:01:16-163840 ERROR Load model: repo="black-forest-labs/FLUX.1-Kontext-dev"
sdnext-ipex | login=True 401 Client Error. (Request ID:
sdnext-ipex | Root=1-6875620c-4ddbac246caa80376513448a;bea8cc0a-
```

```
fb48-
sdnext-ipex | 452c-a738-debd0038cb43)
sdnext-ipex |
sdnext-ipex | Cannot access gated repo for url
sdnext-ipex | https://huggingface.co/api/models/black-forest-
labs/FLU
sdnext-ipex | X.1-Kontext-dev/auth-check.
sdnext-ipex | Access to model black-forest-labs/FLUX.1-Kontext-dev
is
sdnext-ipex | restricted. You must have access to it and be
sdnext-ipex | authenticated to access it. Please log in.
```

## login to huggingface manually

```
# docker exec -it sdnext-openvino bash
# export venv_dir=/mnt/python/venv/
source "${venv_dir}"/bin/activate
git config --global credential.helper store
/mnt/python/venv/bin/huggingface-cli login
# docker restart sdnext-openvino
# Download or use the model after a restart
```

## System info

```
app: sdnext.git updated: 2025-07-14 hash: 79d0fb81 url:
https://github.com/vladmandic/sdnext.git/tree/dev
arch: x86_64 cpu: x86_64 system: Linux release: 6.11.0-29-generic
python: 3.12.3 Torch: 2.7.1+cpu
device: Intel(R) Arc(TM) Graphics (iGPU) openvino: 2025.2.0
ram: free:95.64 used:29.69 total:125.33 gpu: total:125.33
xformers: diffusers: 0.35.0.dev0 transformers: 4.53.2
active: cpu dtype: torch.float32 vae: torch.float32 unet: torch.float32
base: Diffusers/nvidia/Cosmos-Predict2-2B-Text2Image [acdb5fde99] refiner: none vae: none te:
none unet: none

app: sdnext.git updated: 2025-08-02 hash: e90ac68d url:
```

```
https://github.com/vladmandic/sdnext.git/tree/dev
arch: x86_64 cpu: x86_64 system: Linux release: 6.14.0-27-generic
python: 3.12.3 Torch: 2.7.1+cpu
device: Intel(R) Iris(R) Xe Graphics (iGPU) openvino: 2025.2.0
ram: free:59.83 used:34.14 total:93.97 gpu: total:93.97
xformers: diffusers: 0.35.0.dev0 transformers: 4.54.1
active: cpu dtype: torch.float32 vae: torch.float32 unet: torch.float32
base: Diffusers/Kwai-Kolors/Kolors-diffusers [7e091c7519] refiner: none vae: none te: none
UNET: none
```

## openvino (float32) note

Memory consumption in openvino (float32) significantly higher than ipex (bfloat16) so it may cause OOM on same HW with same models than ipex is OK with (Flux.1 dev kontext example, openvino/gpu iGPU Intel Arc)

```

root@server6: ~
Aborted (core dumped)
root@server6:~#

Ultra 9 185H 4.8 GHz
CPU 6% 97°C
C0 0% 78°C C11 0% 78°C
C1 39% 95°C C12 0% 78°C
C2 32% 95°C C13 0% 77°C
C3 22% 69°C C14 0% 78°C
C4 39% 69°C C15 0% 78°C
C5 1% 77°C C16 0% 77°C
C6 1% 78°C C17 0% 78°C
C7 0% 78°C C18 0% 78°C
C8 0% 77°C C19 0% 78°C
C9 0% 78°C C20 0% 95°C
C10 0% 78°C C21 0% 97°C
Load AVG: 4.89 8.30 6.90
GPU 4% 0.15W

total ▲▼ gpu-totals

up 00:42:28

mem disks io
Total: 125 GiB root ▲116K 14.5 TiB
Used: 124 GiB IO%
99% Used: 50% 7.29 TiB
Free: 50% 7.26 TiB
Available: 735 MiB efi 1.04 GiB
1% IO%
Used: 1% 6.14 MiB
Free: 99% 1.04 GiB
Cached: 34.7 GiB
28%
Free: 1.24 GiB
1%

net 46.173.144.30 sync auto zero -b wan n-
30K
download
▼ 688 Byte/s (5.37 Kibps)
▼ Top: (65.6 Mibps)
▼ Total: 42.0 MiB
▲ 19.1 KiB/s (152 Kibps)
▲ Top: (1.12 Mibps)
▲ Total: 48.4 MiB
upload

proc filter per-core reverse tree cpu lazy
Pid: Program: Command: User: MemB Cpu%
26799 python3 /mnt/python/venv/bi root 86G 4.4
26505 kworker/ root 0B 0.0
23345 kworker/ root 0B 0.0
16011 kworker/ root 0B 0.0
15996 kworker/ root 0B 0.0
11478 kworker/ root 0B 0.0
19684 kworker/ root 0B 0.0
23347 kworker/ root 0B 0.0
498 kworker/ root 0B 0.0
26509 kworker/ root 0B 0.0
2381 mysqld mysqld 999 417M 0.0
1294 zabbix_a /usr/sbin/zabbix_ag zabb+ 23M 0.0
180 kswapd0 root 0B 3.0
2465 python3 /usr/local/bin/pyth root 542M 0.0
160 kcompact root 0B 0.0
1366 dockerd /usr/bin/dockerd -H root 55M 0.0
27282 kworker/ root 0B 0.0
21321 btop btop root 6.9M 0.0
4069 python python -m gunicorn root 129M 0.0
23292 kworker/ root 0B 0.0
3194 immich immich root 213M 0.0
2195 valkey-s valkey-server *:637 999 8.2M 0.0
724 kworker/ root 0B 0.0
1297 containe /usr/bin/containerd root 31M 0.0
27281 kworker/ root 0B 0.0
4428 immich-a immich-api root 131M 0.0
104 migratio root 0B 0.0
55 migratio root 0B 0.0
110 migratio root 0B 0.0
2956 openvpn- ./openvpn-ui root 7.5M 0.0
3834 mariadb /usr/bin/mariadb - oliu+ 87M 0.0
3952 python python -m gunicorn root 31M 0.0
3843 apache2 apache2 -DFOREGROUN www++ 77M 0.0
2203 ollama /bin/ollama serve root 11M 0.0
62 migratio root 0B 0.0
4500 memcache memcached -u abc oliu+ 584K 0.0
3841 apache2 apache2 -DFOREGROUN www++ 78M 0.0
3836 apache2 apache2 -DFOREGROUN www++ 79M 0.0
select info terminate kill signals 0/494

```

# Reinstall

## Clean reinstall

Remove all sdnex related date and have clean install

```
docker compose down --rmi local
docker system prune --all --force
rm -Rf /docker/sdnext/*
docker compose up -d
docker compose logs --follow
```

## Preserve downloaded models

save the models and output images

```
docker compose down --rmi local
docker system prune --all --force
rm -Rf /docker/sdnext/app-volume/ /docker/sdnext/python-volume/ /docker/sdnext/mnt-
volume/python/
docker compose up -d
docker compose logs --follow
```

## Config example for iGPU

```
cat /docker/sdnext/mnt-volume/data/config.json
{
  "no_half": false,
  "openvino_devices": [
    "GPU"
  ],
  "samples_filename_pattern": "[seq] - [date] - [model_name] - [height]x[width] - Seed[seed] - CFG[cfg] -
STEP[steps]",
}
```

## Example image generation on iGPU

SD.Next OpenVINO Intel Xe iGPU on Intel core i7-1355U + 96GB DDR5 (35/55W power limits)  
float32

Kolors, Cosmos, Flux, Chroma not work

Model	resolution	time (at 20 STEPS)
-------	------------	--------------------

dreamshaper_8 (SD1.5)	512x512	55.17s
dreamshaper_8 (SD1.5)	768x768	2m 27.75s
cyberrealistic_v90 (SD1.5)	768x768	2m 20.86s
Stable Diffusion 2.1 (SD2.1)	512x512	55.17s
Stable Diffusion 2.1 (SD2.1)	768x768	2m 13.59s
Stable Diffusion 2.1 (SD2.1)	1024x1024	4m 40.88s
TempestV0.1-Artistic (SDXL)	512x512	1m 58.98s
TempestV0.1-Artistic (SDXL)	768x768	3m 23.79s
TempestV0.1-Artistic (SDXL)	1024x1024	5m 24.46s
TempestV0.1-Artistic (SDXL)	1280x1280	8m 25.92s
cyberillustrious_v60Alt (SDXL)	1024x1024	5m 13.36s
cyberrealisticXL_v60 (SDXL)	1024x1024	5m 28.79s
cyberrealisticPony_v127Alt	1024x1024	5m 31.72s
PixArt-Sigma-XL-2-512-MS	512x512	1m 30.74s
PixArt-Sigma-XL-2-1024-MS	1024x1024	6m 26.39s
Stable Diffusion 3.5 medium (SD3.5)	512x512	2m 47.35s
Stable Diffusion 3.5 medium (SD3.5)	768x768	4m 40.50s
Stable Diffusion 3.5 medium (SD3.5)	1024x1024	8m 52.75s
Stable Diffusion 3.5 large (SD3.5)	512x512	7m 7.58s

## Stable Diffusion 3.5 medium



**Prompt:** car

**Parameters:** Steps: 20 | Size: 768x768 | Seed: 116815525 | CFG scale: 6 | App: SD.Next | Version: c0b8c4e | Pipeline: StableDiffusion3Pipeline | Operations: txt2img | Model: stable-diffusion-3.5-medium

Time: 4m 40.50s | total 313.18 pipeline 251.68 preview 32.68 decode 28.80 | RAM 33.87 GB 36%

PixArt-Sigma-XL-2-512-MS



**Prompt:** car

**Parameters:** Steps: 20 | Size: 512x512 | Seed: 2113650425 | CFG scale: 6 | App: SD.Next | Version: a9c65c0 | Pipeline: PixArtSigmaPipeline | Operations: txt2img | Model: PixArt-Sigma-XL-2-512-MS

Time: 1m 30.74s | total 97.25 pipeline 76.88 decode 13.60 preview 6.51 | RAM 22.46 GB 24%

**SD1.5**



**Prompt:** car

**Parameters:** Steps: 20 | Size: 768x768 | Seed: 4104685571 | CFG scale: 6 | App: SD.Next | Version: a9c65c0 | Pipeline: StableDiffusionPipeline | Operations: txt2img | Model: dreamshaper\_8 | Model hash: 879db523c3

Time: 2m 29.85s | total 184.46 pipeline 118.10 preview 34.60 decode 31.73 | RAM 6.66 GB 7%

## Config example for CPU

```

{
  "no_half": false,
  "openvino_devices": [
    "CPU"
  ],
  "samples_filename_pattern": "[seq]-[date]-[model_name]-[height]x[width]-Seed[seed]-CFG[cfg]-STEP[steps]",
  "device_map": "cpu",
  "diffusers_generator_device": "CPU"
}

```

## Example image generation on CPU

SD.Next openvino CPU on Intel core i7-1355U + 96GB DDR5 (25/55W power limits) float32

Model	resolution	time (at 20 STEPS)
Kolors	1280x1280	34m 51.83s
FLUX.1-Krea-dev	512x512	37m 28.47s (96GB RAM + 42GB swap)
FLUX.1-Krea-dev	768x768	96GB RAM + 64GB swap is not enough
dreamshaper_8 (SD1.5)	512x512	2m 58.45s
dreamshaper_8 (SD1.5)	768x768	7m 53.50s
dreamshaper_8 (SD1.5)	1024x1024	18m 10.01s
TempestV0.1-Artistic (SDXL)	512x512	5m 16.53s
TempestV0.1-Artistic (SDXL)	768x768	11m 0.86s
TempestV0.1-Artistic (SDXL)	1024x1024	20m 41.39s
TempestV0.1-Artistic (SDXL)	1280x1280	34m 5.18s

## Kolors



**Prompt:** car

**Parameters:** Steps: 20 | Size: 1280x1280 | Seed: 1747916893 | CFG scale: 6 | App: SD.Next | Version: e90ac68 | Pipeline: KolorsPipeline | Operations: txt2img | Model: Kolors-diffusers

**Execution:** Time: 34m 51.83s | total 2091.82 pipeline 1960.52 decode 131.16 | RAM 90.33 GB 96%



**Prompt:** car

**Parameters:** Steps: 20 | Size: 1024x1024 | Seed: 3229942612 | CFG scale: 6 | App: SD.Next | Version: e90ac68 | Pipeline: KolorsPipeline | Operations: txt2img | Model: Kolors-diffusers

## Flux.1 Krea

93GB of RAM + ~ 46GB of swap



**Prompt:** car

**Parameters:** Steps: 20 | Size: 512x512 | Seed: 1747916893 | CFG scale: 6 | App: SD.Next | Version: e90ac68 | Pipeline: FluxPipeline | Operations: txt2img | Model: FLUX.1-Krea-dev

**Execution:** Time: 37m 28.47s | total 2292.37 pipeline 2224.84 decode 23.42 prompt 19.16 te 19.13 preview 5.59 | RAM 91.66 GB 98%

```

root@server4: ~
07:12:45 PM
- 2000ms +

13th Gen i7-1355U 3.2 GHz
CPU 62% 79°C
C0 89% 66°C C6 67% 66°C
C1 1% 66°C C7 71% 79°C
C2 87% 73°C C8 72% 73°C
C3 0% 73°C C9 83% 73°C
C4 49% 78°C C10 90% 73°C
C5 49% 78°C C11 92% 73°C
Load AVG: 7.36 7.71 7.49
GPU 0% 0.00W

up 05:59:02

mem disks io
Total: 93.9 GiB root 7.21 TiB
Used: 93.1 GiB IO%
99% Used: 2% 145 GiB
Free: 98% 7.07 TiB
Available: 819 MiB swap 63.9 GiB
1% Used: 72% 46.1 GiB
Free: 28% 17.8 GiB
Cached: 441 MiB efi 1.04 GiB
0% IO%
Free: 1.03 GiB Used: 1% 6.14 MiB
Free: 99% 1.04 GiB

net 10.9.9.122 sync auto zero -b enp5s0 n-
77K
download
▼ 1.41 KiB/s (11.3 Kibps)
▼ Top: (52.4 Kibps)
▼ Total: 28.0 GiB
▲ 16.4 KiB/s (131 Kibps)
▲ Top: (2.02 Mibps)
▲ Total: 389 MiB
upload

proc filter per-core reverse tree cpu lazy
Pid: Program: Command: User: MemB Cpu%
2712 python3 python launch.py -f root 92G 62.4
117 kswapd0 root 0B 0.0
16508 kworker/ root 0B 0.0
16509 kworker/ root 0B 0.0
14302 btop btop root 6.0M 0.0
13336 kworker/ root 0B 0.0
885 containe /usr/bin/containerd root 13M 0.0
18243 kworker/ root 0B 0.0
16613 kworker/ root 0B 0.0
98 kcompact root 0B 0.0
1033 dockerd /usr/bin/dockerd -H root 20M 0.0
16926 kworker/ root 0B 0.0
4780 sshd sshd: oliutyi@pts/2 oliu+ 4.2M 0.0
32 migratio root 0B 0.0
18 rcu_pree root 0B 0.0
4790 sudo sudo su - oliu+ 4.3M 0.0
38 migratio root 0B 0.0
462 multipat /sbin/multipathd -d root 26M 0.0
44 migratio root 0B 0.0
2609 containe /usr/bin/containerd root 4.9M 0.0
1 systemd /sbin/init root 8.7M 0.0
56 migratio root 0B 0.0
18505 kworker/ root 0B 0.0
50 migratio root 0B 0.0
2075 fwupd /usr/libexec/fwupd/ root 10M 0.0
26 migratio root 0B 0.0
4066 kworker/ root 0B 0.0
360 btrfs-tr root 0B 0.0
62 migratio root 0B 0.0
21 migratio root 0B 0.0
6424 kworker/ root 0B 0.0
11850 kworker/ root 0B 0.0
837 thermald /usr/sbin/thermald root 6.0M 0.0
2585 kworker/ root 0B 0.0
68 migratio root 0B 0.0
13962 kworker/ root 0B 0.0
685 systemd- /usr/lib/systemd/sy syst+ 6.6M 0.0
141 kworker/ root 0B 0.0
select info terminate kill signals 0/249

```

# Use nginx for SD.Next

## install nginx

```

# SSL certificate
#vi /etc/ssl/private/wildcard.pem

```

```
#chmod 600 /etc/ssl/private/wildcard.pem
#vi /etc/ssl/certs/wildcard.ca
# Nginx
sudo apt install -y nginx
vi /etc/nginx/conf.d/sdnext.conf
systemctl restart nginx
systemctl status nginx
```

## nginx config example

```
server {
    listen      80;
    server_name sdnext.example.com;
    return 301 https://$server_name$request_uri;
}
server {
    listen 443 ssl;
    server_name sdnext.example.com;
    # SSL Settings
    ssl_certificate /etc/ssl/private/wildcard.pem;
    ssl_certificate_key /etc/ssl/private/wildcard.pem;
    ssl_trusted_certificate /etc/ssl/certs/wildcard.ca;
    add_header Strict-Transport-Security 'max-age=15552000; includeSubDomains';
    ssl_protocols TLSv1.3 TLSv1.2;
    ssl_ciphers "EECDH+ECDSA+AESGCM EECDH+aRSA+AESGCM EECDH+ECDSA+SHA384
EECDH+ECDSA+SHA256 EECDH+aRSA+SHA384 EECDH+aRSA+SHA256 EECDH+aRSA+RC4 EECDH EDH+aRSA HIGH !RC4
!aNULL !eNULL !LOW !3DES !MD5 !EXP !PSK !SRP !DSS";
    ssl_session_cache shared:SSL:20m;
    ssl_session_timeout 1h;
    ssl_prefer_server_ciphers on;
    ssl_stapling on;
    ssl_stapling_verify on;
    # Timeouts
    client_max_body_size 50G;
    client_body_timeout 600s;
    proxy_read_timeout 600s;
    proxy_send_timeout 600s;
    send_timeout 600s;
    # Set headers
```

```
proxy_set_header Host          $http_host;
proxy_set_header X-Real-IP     $remote_addr;
proxy_set_header X-Forwarded-For $proxy_add_x_forwarded_for;
proxy_set_header X-Forwarded-Proto $scheme;
# enable websockets
proxy_http_version 1.1;
proxy_set_header Upgrade $http_upgrade;
proxy_set_header Connection "upgrade";
proxy_redirect      off;
# DNS
resolver 1.1.1.1 8.8.4.4 valid=300s;
resolver_timeout 5s;

# ===== Alternative image download page BEGIN
location ~* /(download-text)/.*.(gif|jpg|png)$ {
    add_header Content-disposition "attachment; filename=$1";
    types {}
    default_type application/octet-stream;
    expires -1;
}

location /download-text {
    alias /docker/sdnext/mnt-volume/data/outputs/text/;
    autoindex on;
    expires 1s;
    location ~* ^.+\..(jpeg|gif|png|jpg)
    {
        add_header Content-disposition "attachment";
    }
}
# ===== Alternative image download page END

location / {
    proxy_pass http://127.0.0.1:7860;
}
}
```

# SD.Next btrfs deduplication example

in case SD.Next linux server got btrfs partition we might enjoy deduplication of the files and when we use a large number of models that deduplication may save some space

```
btrfs filesystem du --summarize --human-readable /docker/sdnext/  
compsize -x /docker/sdnext/  
jdupes --one-file-system --dedupe --recurse --size /docker/sdnext/
```

## Execution example

```
root@server6:~ # btrfs filesystem du --summarize --human-readable /docker/sdnext/huggingface-  
volume
```

Total	Exclusive	Set shared	Filename
399.51GiB	399.51GiB	0.00B	/docker/sdnext/huggingface-volume

```
root@server6:~ # compsize -x /docker/sdnext/huggingface-volume
```

Processed 526 files, 1860252 regular extents (1860252 refs), 48 inline.

Type	Perc	Disk Usage	Uncompressed	Referenced
TOTAL	89%	357G	399G	399G
none	100%	175G	175G	175G
zlib	81%	181G	223G	223G

```
root@server6:~ # jdupes --one-file-system --dedupe --recurse --size  
/docker/sdnext/huggingface-volume
```

```
root@server6:~ # btrfs filesystem du --summarize --human-readable /docker/sdnext/huggingface-  
volume
```

Total	Exclusive	Set shared	Filename
399.51GiB	294.70GiB	36.56GiB	/docker/sdnext/huggingface-volume

```
root@server6:~ # compsize -x /docker/sdnext/huggingface-volume
```

Processed 526 files, 1516941 regular extents (1861709 refs), 48 inline.

Type	Perc	Disk Usage	Uncompressed	Referenced
TOTAL	89%	297G	333G	399G

none	100%	151G	151G	175G
zlib	80%	146G	182G	223G

Example run on only one folder. Result might be better on FS level. But just as an example

```

root@server6:~ # jdupes --one-file-system --dedupe --recurse --size
/docker/sdnext/huggingface-volume
Scanning: 353 files, 527 items (in 2 specified)
  [SRC] /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-E1-
Full/blobs/f1ecf4952c2076dc9ef09f54814a2d4b3f14bfa9
  =====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-
Dev/blobs/f1ecf4952c2076dc9ef09f54814a2d4b3f14bfa9
  =====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-
Fast/blobs/f1ecf4952c2076dc9ef09f54814a2d4b3f14bfa9
  =====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-
Full/blobs/f1ecf4952c2076dc9ef09f54814a2d4b3f14bfa9

  [SRC] /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-
Dev/blobs/cf6c4d78b3b4f575a8b296ee5fccf6ab4e14c686
  =====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-
Fast/blobs/cf6c4d78b3b4f575a8b296ee5fccf6ab4e14c686
  =====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-
Full/blobs/cf6c4d78b3b4f575a8b296ee5fccf6ab4e14c686

  [SRC] /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-
Dev/blobs/f99e44fb320546035ca153638f1f2276e0cd2e04
  =====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-
Fast/blobs/f99e44fb320546035ca153638f1f2276e0cd2e04
  =====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-
Full/blobs/f99e44fb320546035ca153638f1f2276e0cd2e04

  [SRC] /docker/sdnext/huggingface-volume/hub/models--Wan-AI--Wan2.1-T2V-1.3B-
Diffusers/blobs/481c7b2b39771c44df6dd8d13ee12ed072d731b4a650bd092885d4d52db229ad
  =====> /docker/sdnext/huggingface-volume/hub/models--Wan-AI--Wan2.1-T2V-14B-
Diffusers/blobs/481c7b2b39771c44df6dd8d13ee12ed072d731b4a650bd092885d4d52db229ad

  [SRC] /docker/sdnext/huggingface-volume/hub/models--Wan-AI--Wan2.1-T2V-1.3B-
Diffusers/blobs/c0ef3a140898e228a3520c9adec60743d2e8e5b3d229651bb37f1a3921919f99
  =====> /docker/sdnext/huggingface-volume/hub/models--Wan-AI--Wan2.1-T2V-14B-

```

Diffusers/blobs/c0ef3a140898e228a3520c9adec60743d2e8e5b3d229651bb37f1a3921919f99

[SRC] /docker/sdnext/huggingface-volume/hub/models--Wan-AI--Wan2.1-T2V-1.3B-Diffusers/blobs/a451792c739c05bca4606190cc2dd16731411bac03b4cf6aacc5767321f857c9  
====> /docker/sdnext/huggingface-volume/hub/models--Wan-AI--Wan2.1-T2V-14B-Diffusers/blobs/a451792c739c05bca4606190cc2dd16731411bac03b4cf6aacc5767321f857c9

[SRC] /docker/sdnext/huggingface-volume/hub/models--Wan-AI--Wan2.1-T2V-1.3B-Diffusers/blobs/f93148bcc04052a169e1e49bfcf6125df6cf9bf243cb9c627da75266cf8e35c3  
====> /docker/sdnext/huggingface-volume/hub/models--Wan-AI--Wan2.1-T2V-14B-Diffusers/blobs/f93148bcc04052a169e1e49bfcf6125df6cf9bf243cb9c627da75266cf8e35c3

[SRC] /docker/sdnext/huggingface-volume/hub/models--Wan-AI--Wan2.1-T2V-1.3B-Diffusers/blobs/7e76e18d224531b8197a46231cb53daf7f2f6ca707130252becf933026ac4eea  
====> /docker/sdnext/huggingface-volume/hub/models--Wan-AI--Wan2.1-T2V-14B-Diffusers/blobs/7e76e18d224531b8197a46231cb53daf7f2f6ca707130252becf933026ac4eea

[SRC] /docker/sdnext/huggingface-volume/hub/models--Wan-AI--Wan2.1-T2V-1.3B-Diffusers/blobs/60ece61b46ecb3e6a5b705ea304bc97535317c2a  
====> /docker/sdnext/huggingface-volume/hub/models--Wan-AI--Wan2.1-T2V-14B-Diffusers/blobs/60ece61b46ecb3e6a5b705ea304bc97535317c2a

[SRC] /docker/sdnext/huggingface-volume/hub/models--Wan-AI--Wan2.1-T2V-1.3B-Diffusers/blobs/2fd01c57ddd8ae386d518c69c087c8ba8c73804  
====> /docker/sdnext/huggingface-volume/hub/models--Wan-AI--Wan2.1-T2V-14B-Diffusers/blobs/2fd01c57ddd8ae386d518c69c087c8ba8c73804

[SRC] /docker/sdnext/huggingface-volume/hub/models--Freepik--F-Lite/blobs/ec87bffd1923e8b2774a6d240c922a41f6143081d52cf83b8fe39e9d838c893e  
====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-Dev/blobs/ec87bffd1923e8b2774a6d240c922a41f6143081d52cf83b8fe39e9d838c893e  
====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-Fast/blobs/ec87bffd1923e8b2774a6d240c922a41f6143081d52cf83b8fe39e9d838c893e  
====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-Full/blobs/ec87bffd1923e8b2774a6d240c922a41f6143081d52cf83b8fe39e9d838c893e  
====> /docker/sdnext/huggingface-volume/hub/models--ostris--Flex.2-preview/blobs/ec87bffd1923e8b2774a6d240c922a41f6143081d52cf83b8fe39e9d838c893e

[SRC] /docker/sdnext/huggingface-volume/hub/models--Freepik--F-Lite/blobs/a5640855b301fcdcbceddfa90ae8066cd9414aff020552a201a255ecf2059da00

```
====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-
Dev/blobs/a5640855b301fcdbceddfa90ae8066cd9414aff020552a201a255ecf2059da00
====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-
Fast/blobs/a5640855b301fcdbceddfa90ae8066cd9414aff020552a201a255ecf2059da00
====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-
Full/blobs/a5640855b301fcdbceddfa90ae8066cd9414aff020552a201a255ecf2059da00
====> /docker/sdnext/huggingface-volume/hub/models--ostris--Flex.2-
preview/blobs/a5640855b301fcdbceddfa90ae8066cd9414aff020552a201a255ecf2059da00
```

```
[SRC] /docker/sdnext/huggingface-volume/hub/models--Freepik--F-
Lite/blobs/c8728bc3dca59d2616a2a594cbac3ddb9eb77d5b
====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-
Dev/blobs/c8728bc3dca59d2616a2a594cbac3ddb9eb77d5b
====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-
Fast/blobs/c8728bc3dca59d2616a2a594cbac3ddb9eb77d5b
====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-
Full/blobs/c8728bc3dca59d2616a2a594cbac3ddb9eb77d5b
====> /docker/sdnext/huggingface-volume/hub/models--ostris--Flex.2-
preview/blobs/c8728bc3dca59d2616a2a594cbac3ddb9eb77d5b
```

```
[SRC] /docker/sdnext/huggingface-volume/hub/models--nvidia--Cosmos-Predict2-2B-
Text2Image/blobs/30dea2006604226b944972536b9fc1ae40e3ef39b844140641cc450af5490158
====> /docker/sdnext/huggingface-volume/hub/models--nvidia--Cosmos-Predict2-14B-
Text2Image/blobs/30dea2006604226b944972536b9fc1ae40e3ef39b844140641cc450af5490158
```

```
[SRC] /docker/sdnext/huggingface-volume/hub/models--nvidia--Cosmos-Predict2-2B-
Text2Image/blobs/d90c625c136567ef62dc13d7fde0d54393a2b08e7400b3594d38c6a9ba12f897
====> /docker/sdnext/huggingface-volume/hub/models--nvidia--Cosmos-Predict2-14B-
Text2Image/blobs/d90c625c136567ef62dc13d7fde0d54393a2b08e7400b3594d38c6a9ba12f897
```

```
[SRC] /docker/sdnext/huggingface-volume/hub/models--nvidia--Cosmos-Predict2-2B-
Text2Image/blobs/371ddccda7716967086fd98ef7e2f8e14f96f6ae
====> /docker/sdnext/huggingface-volume/hub/models--nvidia--Cosmos-Predict2-14B-
Text2Image/blobs/371ddccda7716967086fd98ef7e2f8e14f96f6ae
```

```
[SRC] /docker/sdnext/huggingface-volume/hub/models--nvidia--Cosmos-Predict2-2B-
Text2Image/blobs/6c85dc465ff1a4392f21aefb7ab6ea7b9b990a9
====> /docker/sdnext/huggingface-volume/hub/models--nvidia--Cosmos-Predict2-14B-
Text2Image/blobs/6c85dc465ff1a4392f21aefb7ab6ea7b9b990a9
```

on /docker/sdnext level it looks like

```
[SRC] /docker/sdnext/huggingface-volume/hub/models--Freepik--F-
Lite/blobs/ec87bffd1923e8b2774a6d240c922a41f6143081d52cf83b8fe39e9d838c893e
====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-
Dev/blobs/ec87bffd1923e8b2774a6d240c922a41f6143081d52cf83b8fe39e9d838c893e
====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-
Fast/blobs/ec87bffd1923e8b2774a6d240c922a41f6143081d52cf83b8fe39e9d838c893e
====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-
Full/blobs/ec87bffd1923e8b2774a6d240c922a41f6143081d52cf83b8fe39e9d838c893e
====> /docker/sdnext/huggingface-volume/hub/models--ostris--Flex.2-
preview/blobs/ec87bffd1923e8b2774a6d240c922a41f6143081d52cf83b8fe39e9d838c893e
====> /docker/sdnext/mnt-volume/models/Diffusers/models--black-forest-labs--FLUX.1-
dev/blobs/ec87bffd1923e8b2774a6d240c922a41f6143081d52cf83b8fe39e9d838c893e
====> /docker/sdnext/mnt-volume/models/Diffusers/models--black-forest-labs--FLUX.1-Kontext-
dev/blobs/ec87bffd1923e8b2774a6d240c922a41f6143081d52cf83b8fe39e9d838c893e
====> /docker/sdnext/mnt-volume/models/Diffusers/models--black-forest-labs--FLUX.1-
schnell/blobs/ec87bffd1923e8b2774a6d240c922a41f6143081d52cf83b8fe39e9d838c893e
====> /docker/sdnext/mnt-volume/models/Diffusers/models--lodestones--
Chroma/blobs/ec87bffd1923e8b2774a6d240c922a41f6143081d52cf83b8fe39e9d838c893e
====> /docker/sdnext/mnt-volume/models/Diffusers/models--ostris--Flex.2-
preview/blobs/ec87bffd1923e8b2774a6d240c922a41f6143081d52cf83b8fe39e9d838c893e
====> /docker/sdnext/mnt-volume/models/Diffusers/models--shuttleai--shuttle-3.1-
aesthetic/blobs/ec87bffd1923e8b2774a6d240c922a41f6143081d52cf83b8fe39e9d838c893e
```

```
[SRC] /docker/sdnext/huggingface-volume/hub/models--Freepik--F-
Lite/blobs/a5640855b301fcdcbceddfa90ae8066cd9414aff020552a201a255ecf2059da00
====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-
Dev/blobs/a5640855b301fcdcbceddfa90ae8066cd9414aff020552a201a255ecf2059da00
====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-
Fast/blobs/a5640855b301fcdcbceddfa90ae8066cd9414aff020552a201a255ecf2059da00
====> /docker/sdnext/huggingface-volume/hub/models--HiDream-ai--HiDream-I1-
Full/blobs/a5640855b301fcdcbceddfa90ae8066cd9414aff020552a201a255ecf2059da00
====> /docker/sdnext/huggingface-volume/hub/models--ostris--Flex.2-
preview/blobs/a5640855b301fcdcbceddfa90ae8066cd9414aff020552a201a255ecf2059da00
====> /docker/sdnext/mnt-volume/models/Diffusers/models--black-forest-labs--FLUX.1-
dev/blobs/a5640855b301fcdcbceddfa90ae8066cd9414aff020552a201a255ecf2059da00
====> /docker/sdnext/mnt-volume/models/Diffusers/models--black-forest-labs--FLUX.1-Kontext-
dev/blobs/a5640855b301fcdcbceddfa90ae8066cd9414aff020552a201a255ecf2059da00
====> /docker/sdnext/mnt-volume/models/Diffusers/models--black-forest-labs--FLUX.1-
```

```
schnell/blobs/a5640855b301fcdbceddfa90ae8066cd9414aff020552a201a255ecf2059da00
====> /docker/sdnext/mnt-volume/models/Diffusers/models--lodestones--
Chroma/blobs/a5640855b301fcdbceddfa90ae8066cd9414aff020552a201a255ecf2059da00
====> /docker/sdnext/mnt-volume/models/Diffusers/models--ostris--Flex.2-
preview/blobs/a5640855b301fcdbceddfa90ae8066cd9414aff020552a201a255ecf2059da00
====> /docker/sdnext/mnt-volume/models/Diffusers/models--shuttleai--shuttle-3.1-
aesthetic/blobs/a5640855b301fcdbceddfa90ae8066cd9414aff020552a201a255ecf2059da00
...
Deduplication done (6565 files processed)
```

# Edit SD.Next task que using sqlite3

Sometimes it is not possible to remove hanged task from SD.Next - it is instantly re-scheduled.

that is can be done by direct edit of sqlite3 database like shown below

```
# apt install sqlite3
sqlite3 /docker/sdnext/app-volume/extensions-builtin/sd-webui-agent-
scheduler/task_scheduler.sqlite3
```

```
sqlite> .tables
app_state  task
```

```
sqlite> .schema task
```

```
CREATE TABLE task (
    id VARCHAR(64) NOT NULL,
    api_task_id VARCHAR(64),
    api_task_callback VARCHAR(255),
    name VARCHAR(255),
    type VARCHAR(20) NOT NULL,
    params TEXT NOT NULL,
    script_params BLOB NOT NULL,
    priority INTEGER NOT NULL,
    status VARCHAR(20) NOT NULL,
    result TEXT,
    bookmarked BOOLEAN,
    created_at DATETIME DEFAULT (datetime('now')) NOT NULL,
    updated_at DATETIME DEFAULT (datetime('now')) NOT NULL,
    PRIMARY KEY (id)
);
```

```
sqlite> .schema task
```

```
CREATE TABLE task (
    id VARCHAR(64) NOT NULL,
    api_task_id VARCHAR(64),
```

```
api_task_callback VARCHAR(255),
name VARCHAR(255),
type VARCHAR(20) NOT NULL,
params TEXT NOT NULL,
script_params BLOB NOT NULL,
priority INTEGER NOT NULL,
status VARCHAR(20) NOT NULL,
result TEXT,
bookmarked BOOLEAN,
created_at DATETIME DEFAULT (datetime('now')) NOT NULL,
updated_at DATETIME DEFAULT (datetime('now')) NOT NULL,
PRIMARY KEY (id)
);
sqlite> SELECT id, name, status FROM task;
task(fhraih0ryzwmliq)||done
task(93jw8bgegiaczkr)||pending
220fd698-05db-4224-8480-03ced279d906||pending
sqlite> DELETE FROM task WHERE id = '220fd698-05db-4224-8480-03ced279d906';

sqlite> .exit
```

# SD.Next by liutyi docker compose

Personal preferences fork from [Vlad SD.Next](#) for Intel Arc iGPU system based on Intel Core Ultra products. Might be also compatible with All other HW that SD.Next is suppose to work on (but that is upstream project is better to be used). Fork is also

1. keeps **standard** theme as default.
2. keeps **scheduler** as built in plugin.
3. uses **long filenames** by default.
4. Have Dockerfile and [containers](#) that based on **consumer Intel drivers** not DC product one.
5. Keeps Intel Extension for Pytorch (ipex) compatibility (as long as it is relevant).

So there are 2 versions available

## IPEX

- Container: <https://hub.docker.com/repository/docker/liutyi/sdnext-ipex/tags/25.11-ipex/>
- Source code: <https://github.com/liutyi/sdnext/tree/ipex>

PyTorch 2.7.1+xpu; IPEX 2.7.10+xpu. It was working better than Pytorch 2.7.1+xpu w/o ipex for iGPU. But might not be actually that better than 2.9.1+xpu. The only reason (at least it seems to) to continue use it is that it provides GPU memory usage verbose stats.

## PyTorch

- Docker container: <https://hub.docker.com/repository/docker/liutyi/sdnext-ipex/tags/25.11-pytorch>
- Source code: <https://github.com/liutyi/sdnext/tree/pytorch>

Based on PyTorch 2.9.1+xpu

## Docker-compose example

```
services:
  sdnnext-ipex:
    container_name: sdnnext-ipex
    # PyTorch 2.7.1+xpu and ipex 2.7.10+xpu for Intel
#   image: liutyi/sdnnext-ipex:25.11-ipex
    # PyTorch 2.9.1+xpu for Intel
    image: liutyi/sdnnext-ipex:25.11-pytorch
    restart: unless-stopped
    devices:
      # GPU
      - /dev/dri:/dev/dri
      # NPU (optional)
      - /dev/accel:/dev/accel
    environment:
      - HF_DEBUG=True
      # To USE Nano Banana get key https://aistudio.google.com/api-keys
#   - GOOGLE_API_KEY=AIxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
    ports:
      - 7860:7860
    volumes:
      - /docker/sdnnext/app-volume:/app
      - /docker/sdnnext/mnt-volume:/mnt
      - /docker/sdnnext/hf-volume:/root/.cache/huggingface
      - /dev/shm:/dev/shm
```