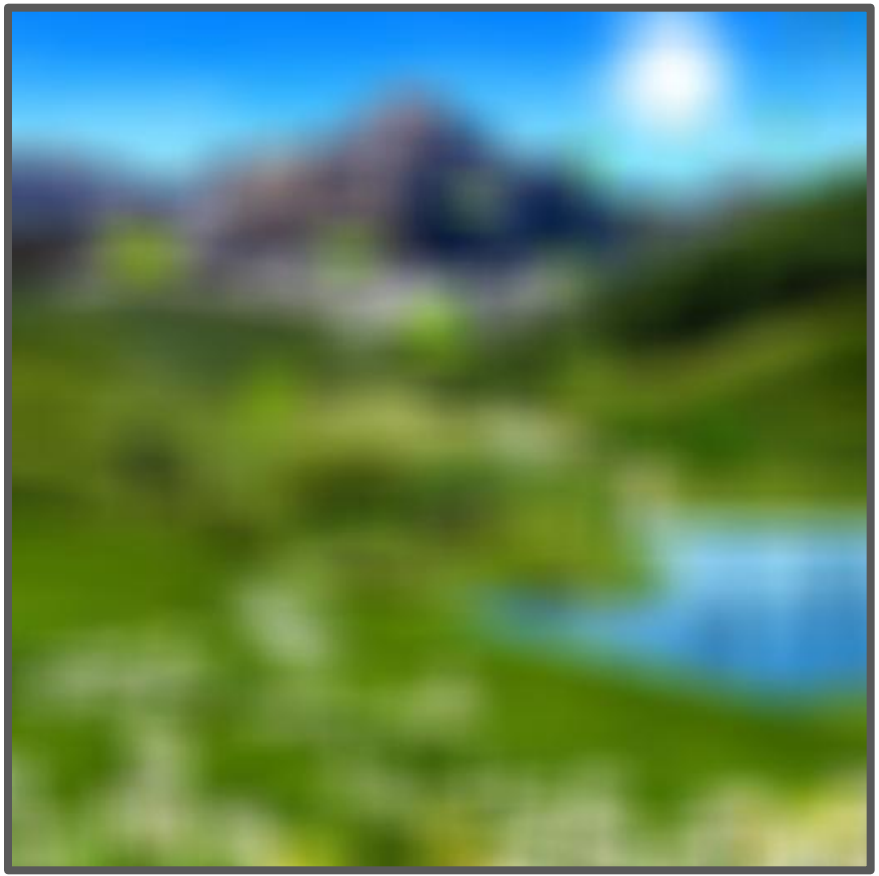
Gaussian Blur

Gaussian Blur

- For our project we decided to implement the Gaussian Blur function

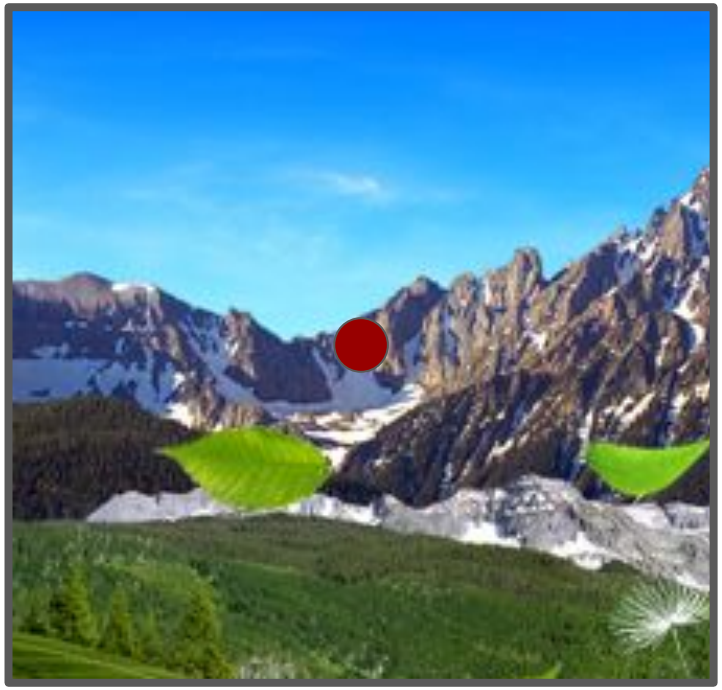
Gaussian Blur

- This function produces an image with a blur filter applied to it



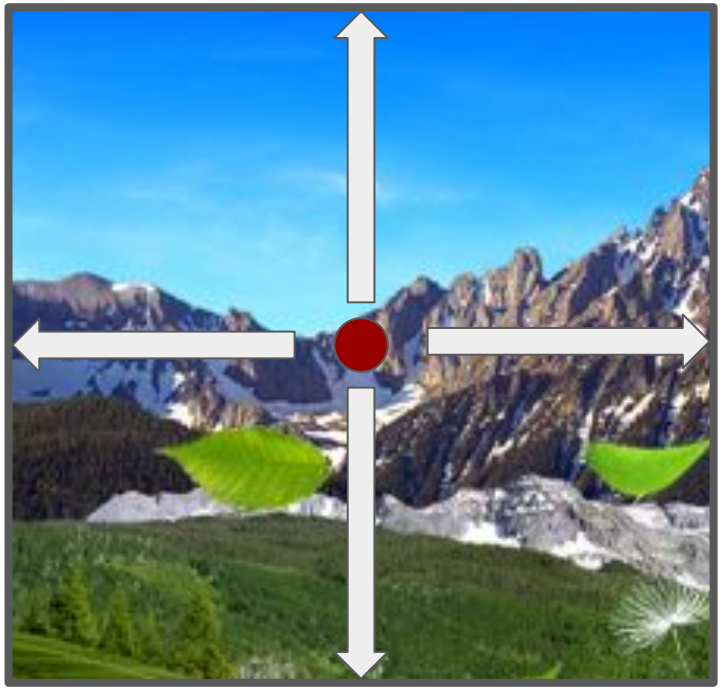
Gaussian Blur

- This involves iterating through all the pixels of the input image.



Gaussian Blur

- Then we further iterate all the pixels within the radius of the current pixel.



Gaussian Blur

- If a pixel is near a corner or an edge, the radius for that side will shrink to a size that will not exceed said edge or corner

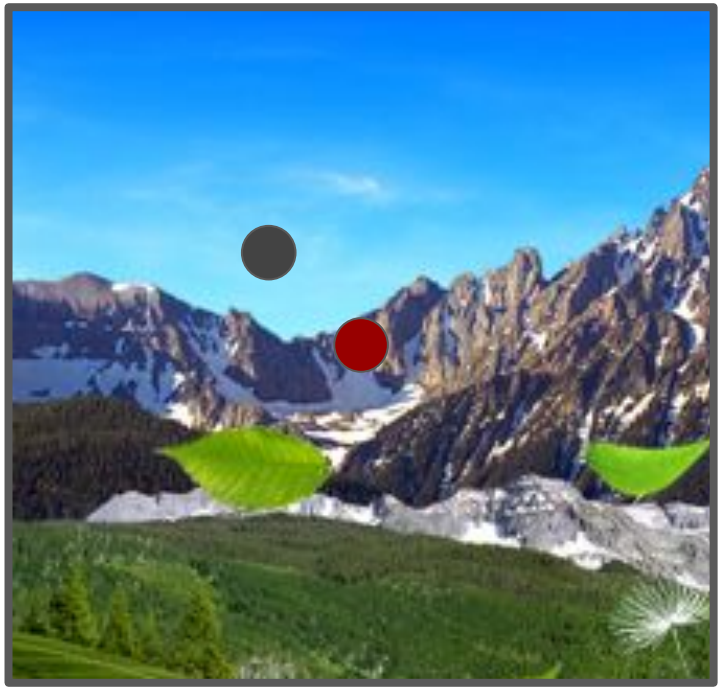
- It will only iterate through this new range and this results in processing fewer pixels than if we are to use the full radius.

Gaussian Blur

- Then, for each pixel R within the radius of the current pixel, we first get its weight via the formula below.

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

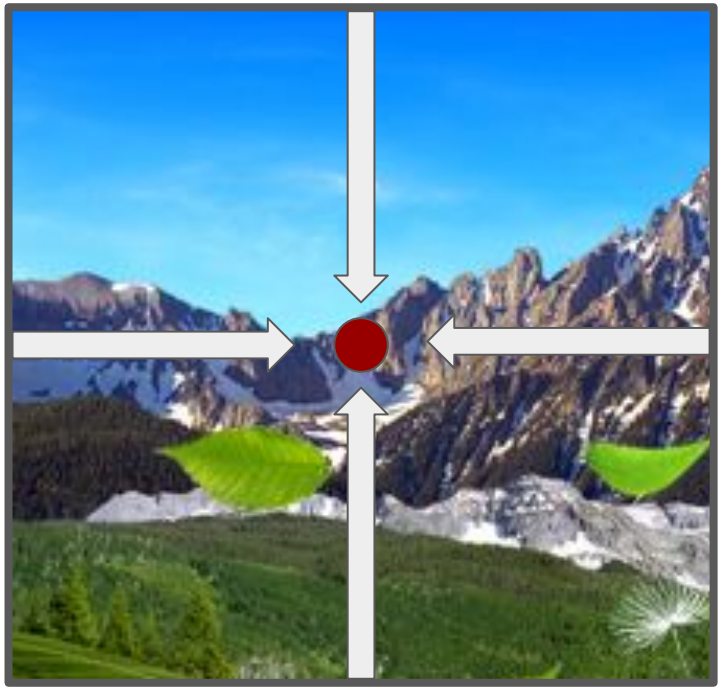
- Afterwards we multiply the received weight to the color of the pixel R



Gaussian Blur

- We then add in all the resulting colors from all the pixels within the Radius of our current pixel.

- After which is it is normalized by dividing the resulting sum with the maximum value you can get.



Gaussian Blur

- This is then the new color of the center pixel
- Once we do this for all the pixels of an image, this will result in an image with a blur filter applied

CUDA Program Parallel Algorithms Implementation

Parallel Algorithm

- In the sequential implementation of the Gaussian Blur, each pixel iterates to obtain the output image and for each output pixel, its neighborhood, radius and standard deviation are iterated to calculate the weighted average color value.
- The CUDA program implements a parallel algorithm for Gaussian Blur, contained within the __global__ function, GaussianBlur

```
std::cout << "Will Process Image "<< std::endl;
for(int i = 0; i < 10; i++) {
  GaussianBlur<<<blocks, threads>>>(gaussianParams);
}
```

```
__global__
void GaussianBlur(GaussianParams *g_param){
```

Parallel Algorithm

```
int index = blockIdx.x * blockDim.x + threadIdx.x;
int stride = blockDim.x * gridDim.x;
int pixelCount = g_param->width * g_param->height;

for (int i = index; i < pixelCount; i += stride){
```

CUDA Kernel

(Base Grid-Stride Loop Implementation)

- What is GSL (Grid-Stride Loop)?
- Why use GSL for Gaussian Blur?
- Overview of baseline implementation

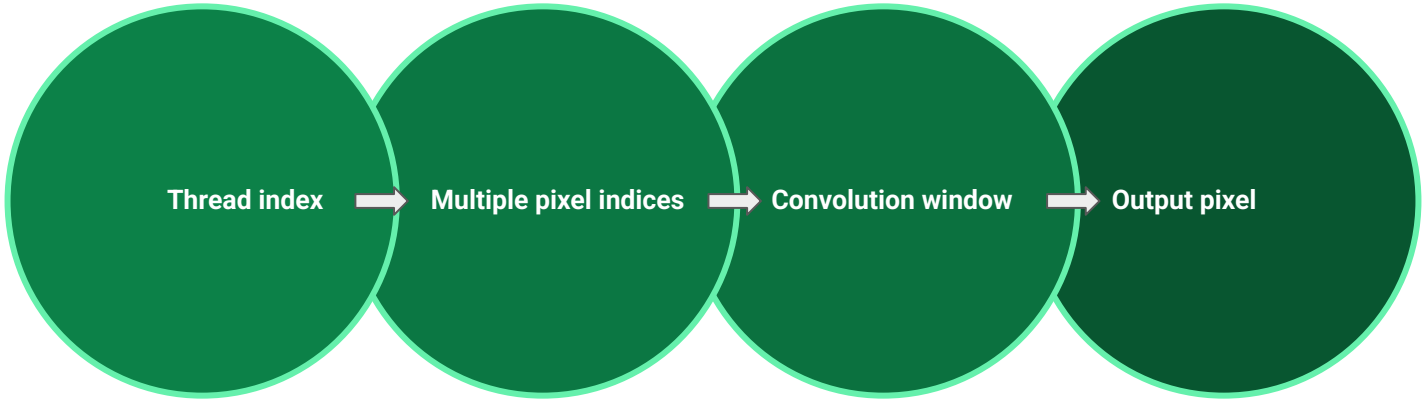
What Base GSL is Doing?

(Base GSL Implementation)

- Each thread processes multiple pixels using grid-stride loop
- No shared memory optimization
- Direct global memory loads
- For each pixel: compute window → accumulate → normalize
- Works for any image size
- Maximizes GPU occupancy
- No need to tune block count per image dimension
- Each thread handles multiple pixels

Kernel Overview Diagram

(Base GSL Implementation)



Core of the GSL

(Base GSL Implementation)

```
int index = blockIdx.x * blockDim.x + threadIdx.x;
int stride = blockDim.x * gridDim.x;
int pixelCount = g_param->width * g_param->height;

for (int i = index; i < pixelCount; i += stride){
```

Gaussian Weight Computation

(Base GSL Implementation)

```
float e_pow = (x*x + y*y) / (2.0f * std_dev * std_dev) * -1.0f;
float e = expf(e_pow);
float frac = 1.0f / sqrtf(2.0f * M_PIf * (std_dev * std_dev));

return frac * e;
```

Pixel Accumulation Loop

(Base GSL Implementation)

```
for (int y = minY; y < maxY; y++){
  for (int x = minX; x < maxX; x++){
    float weight = ComputeWeight(curr_x, curr_y, (float)x, (float)y, g_param->std_dev);
    int idx = y * g_param->width + x;

    totR += g_param->in_r[idx] * weight;
    totG += g_param->in_g[idx] * weight;
    totB += g_param->in_b[idx] * weight;

    baseMax += weight;
```

Boundary Handling

(Base GSL Implementation)

```
int minX = max(curr_x - g_param->radius, 0);
int maxX = min(curr_x + g_param->radius + 1, g_param->width);

int minY = max(curr_y - g_param->radius, 0);
int maxY = min(curr_y + g_param->radius + 1, g_param->height);
```

Normalization of Output Pixel

(Base GSL Implementation)

```
if (baseMax > 1e-6f) {
  float ceilVal = 1.0f / baseMax;
  g_param->out_r[i] = totR * ceilVal;
  g_param->out_g[i] = totG * ceilVal;
  g_param->out_b[i] = totB * ceilVal;
} else {
  g_param->out_r[i] = g_param->in_r[i];
  g_param->out_g[i] = g_param->in_g[i];
  g_param->out_b[i] = g_param->in_b[i];
```

RMSE Validation

(Base GSL Implementation)

```
cv::Mat compImage = cv::imread(compPath, cv::IMREAD_COLOR);
cv::Mat outputImage = cv::imread("512x512_outputImage.jpg", cv::IMREAD_COLOR);

std::cout << "Validation from image" << std::endl;
auto rmse = RSME(gaussianParams, compImage);
std::cout << "RMSE from image: " << rmse << std::endl;

std::cout << "Validation from file" << std::endl;
auto rmse2 = RSME_FromFile(gaussianParams, "output512.csv");
std::cout << "RMSE from file " << rmse2 << std::endl;
```

CUDA Kernel

(GSL + Prefetch)

- This implementation introduces the use of Data Prefetching
- The features are the same as those in the GSL implementation except for the use of cudaMemPrefetchAsync

```
int device = -1;
cudaGetDevice(&device);
```

```
// Prefetch data from CPU-GPU
cudaMemPrefetchAsync(gaussianParams->in_r,array_byte_size, device, NULL);
cudaMemPrefetchAsync(gaussianParams->in_g,array_byte_size, device, NULL);
cudaMemPrefetchAsync(gaussianParams->in_b,array_byte_size, device, NULL);
```

```
// Prefetch data from GPU-CPU
cudaMemPrefetchAsync(gaussianParams->out_r,array_byte_size, cudaCpuDeviceId, NULL);
cudaMemPrefetchAsync(gaussianParams->out_g,array_byte_size, cudaCpuDeviceId, NULL);
cudaMemPrefetchAsync(gaussianParams->out_b,array_byte_size, cudaCpuDeviceId, NULL);
```

Core of the GSL + Prefetch

(GSL + Prefetch Implementation)

```
// Prefetch data from CPU-GPU
cudaMemPrefetchAsync(gaussianParams->in_r,array_byte_size, device, NULL);
cudaMemPrefetchAsync(gaussianParams->in_g,array_byte_size, device, NULL);
cudaMemPrefetchAsync(gaussianParams->in_b,array_byte_size, device, NULL);
```

```
// Prefetch data from GPU-CPU
cudaMemPrefetchAsync(gaussianParams->out_r,array_byte_size, cudaCpuDeviceId, NULL);
cudaMemPrefetchAsync(gaussianParams->out_g,array_byte_size, cudaCpuDeviceId, NULL);
cudaMemPrefetchAsync(gaussianParams->out_b,array_byte_size, cudaCpuDeviceId, NULL);
```

CUDA Kernel

(GSL + Prefetch + Page Creation)

- With Page Creation, we are reducing the CPU and GPU page faults by prefetching blank data from both the CPU and GPU page memories.
- This implementation appends the GSL + Prefetch Kernel.
- Similar to Prefetch, it uses the cuda instruction: cudaMemPrefetchAsync
- We perform this to our parameters in the gaussianParams class object.

```
// Page Creation
//"prefetch data" to create CPU page memory
cudaMemPrefetchAsync(gaussianParams->in_r,array_byte_size, cudaCpuDeviceId, NULL);
cudaMemPrefetchAsync(gaussianParams->in_g,array_byte_size, cudaCpuDeviceId, NULL);
cudaMemPrefetchAsync(gaussianParams->in_b,array_byte_size, cudaCpuDeviceId, NULL);

//"prefetch data" to create GPU page memory
cudaMemPrefetchAsync(gaussianParams->out_r,array_byte_size, device, NULL);
cudaMemPrefetchAsync(gaussianParams->out_g,array_byte_size, device, NULL);
cudaMemPrefetchAsync(gaussianParams->out_b,array_byte_size, device, NULL);
```

CUDA Kernel

(GSL + Prefetch + Page Creation + Memory Advise)

- With Memory Advise, we are further reducing the CPU and GPU page faults by advising Cuda with the memory pointer.
- This implementation appends the GSL + Prefetch + Page Creation Kernel.
- It uses the cuda instruction: cudaMemAdvise
- We perform this to our “in” parameters in the gaussianParams class object.

```
// Memory Advise
 cudaMemAdvise(gaussianParams->in_r, array_byte_size, cudaMemAdviseSetPreferredLocation, cudaCpuDeviceId);
 cudaMemAdvise(gaussianParams->in_r, array_byte_size, cudaMemAdviseSetReadMostly, cudaCpuDeviceId);
 cudaMemAdvise(gaussianParams->in_g, array_byte_size, cudaMemAdviseSetPreferredLocation, cudaCpuDeviceId);
 cudaMemAdvise(gaussianParams->in_g, array_byte_size, cudaMemAdviseSetReadMostly, cudaCpuDeviceId);
 cudaMemAdvise(gaussianParams->in_b, array_byte_size, cudaMemAdviseSetPreferredLocation, cudaCpuDeviceId);
 cudaMemAdvise(gaussianParams->in_b, array_byte_size, cudaMemAdviseSetReadMostly, cudaCpuDeviceId);
```

Execution Time Comparison

Sequential vs Parallel (512x512 px)

	Sequential	Parallel				
	C++	GSL	GSL + Prefetch	GSL + Prefetch + Page Creation	GSL + Prefetch + Page Creation + Mem Advise	Speed Up
Average run time (10 calls)	11,955.1 ms	7.7283ms	7.6563ms	7.6603ms	7.6607ms	
Host to Device		0.4424540 ms	0.3583930m s	0.273113ms	0.272026ms	
Device to Host		0.3033850 ms	0.2440280m s	0.247385ms	0.243802ms	
GPU Kernel Time		8.474139m s	8.258721ms	8.180798ms	8.176528ms	1,462.12

Execution Time Comparison

Sequential vs Parallel (1024x1024 px)

	Sequential	Parallel				
	C++	GSL	GSL + Prefetch	GSL + Prefetch + Page Creation	GSL + Prefetch + Page Creation + Mem Advise	Speed Up
Average run time (10 calls)	114,801 ms	22.636ms	24.260ms	23.850ms	23.524ms	
Host to Device		1.235905ms	1.060581ms	1.05895ms	1.049413ms	
Device to Host		1.104040ms	0.9667900 ms	0.965322ms	0.962281ms	
GPU Kernel Time		24.975945ms	26.287371 ms	25.874272ms	25.535694ms	4,495.707068

Execution Time Comparison

Sequential vs Parallel (2048x2048 px)

	Sequential	Parallel				
	C++	GSL	GSL + Prefetch	GSL + Prefetch + Page Creation	GSL + Prefetch + Page Creation + Mem Advise	Speed Up
Average run time (10 calls)	202,619 ms	73.874ms	73.701ms	70.213ms	71.632ms	
Host to Device		4.983810ms	4.192919ms	4.193435ms	4.182204ms	
Device to Host		4.291367ms	3.856018ms	3.854532ms	3.848360ms	
GPU Kernel Time		83.149177ms	81.749937 ms	78.260967ms	79.653564ms	2,543.753096

Validation Comparison

Validation from Image

		GSL	GSL + Prefetch	GSL + Prefetch + Page Creation	GSL + Prefetch + Page Creation + Mem Advise
512x512	MSE	2.3412e-05	2.3412e-05	2.3412e-05	2.3412e-05
	RMSE	0.00483859	0.00483859	0.00483859	0.00483859
1024x1024	MSE	1.5554e-05	1.5554e-05	1.5554e-05	1.5554e-05
	RMSE	0.00394385	0.00394385	0.00394385	0.00394385
2048x2048	MSE	1.35444e-05	1.35444e-05	1.35444e-05	1.35444e-05
	RMSE	0.00368026	0.00368026	0.00368026	0.00368026

Validation Comparison

Validation from File

		GSL	GSL + Prefetch	GSL + Prefetch + Page Creation	GSL + Prefetch + Page Creation + Mem Advise
512x512	MSE	1.06483e-05	1.06483e-05	1.06483e-05	1.06483e-05
	RMSE	0.00326317	0.00326317	0.00326317	0.00326317
1024x1024	MSE	4.27375e-06	4.27375e-06	4.27375e-06	4.27375e-06
	RMSE	0.00206731	0.00206731	0.00206731	0.00206731
2048x2048	MSE	2.09942e-06	2.09942e-06	2.09942e-06	2.09942e-06
	RMSE	0.00144894	0.00144894	0.00144894	0.00144894

nvprof Data 512x512

512 GSL

```
!nvprof ./imageLoad
```

```
==27095== NVPROF is profiling process 27095, command: ./imageLoad
Image loaded successfully!
Image dimensions: 512x512
Will Process Image
Finished processing.
512x512_outputImage.jpg saved.
Validation from image
MSE: 2.3412e-05
RMSE from image: 0.00483859
Validation from file
MSE: 1.06483e-05
RMSE from file 0.00326317
==27095== Profiling application: ./imageLoad
==27095== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:  100.00%  77.234ms        10  7.7234ms  7.6210ms  8.5657ms  GaussianBlur(GaussianParams*)
      API calls:   58.59%  110.47ms         7  15.781ms  2.4390us  110.38ms  cudaMallocManaged
                   40.92%  77.147ms         1  77.147ms  77.147ms  77.147ms  cudaDeviceSynchronize
                    0.25%  475.17us         7  67.881us  10.530us  105.33us  cudaFree
                    0.15%  288.24us        10  28.823us  5.0800us  188.20us  cudaLaunchKernel
                    0.08%  146.38us       114  1.2840us     102ns  63.660us  cuDeviceGetAttribute
                    0.01%  12.165us         1  12.165us  12.165us  12.165us  cuDeviceGetName
                    0.00%  5.4110us         1  5.4110us  5.4110us  5.4110us  cuDeviceGetPCIBusId
                    0.00%  1.4640us         3     488ns     135ns  1.0790us  cuDeviceGetCount
                    0.00%     841ns         2     420ns     206ns     635ns  cuDeviceGet
                    0.00%     600ns         1     600ns     600ns     600ns  cuModuleGetLoadingMode
                    0.00%     587ns         1     587ns     587ns     587ns  cuDeviceTotalMem
                    0.00%     465ns         1     465ns     465ns     465ns  cudaGetLastError
                    0.00%     216ns         1     216ns     216ns     216ns  cuDeviceGetUuid

==27095== Unified Memory profiling result:
Device "Tesla T4 (0)"
   Count  Avg Size  Min Size  Max Size  Total Size  Total Time  Name
      37  110.70KB  4.0000KB  0.9961MB  4.000000MB  441.0130us  Host To Device
      31  101.16KB  4.0000KB  512.00KB  3.062500MB  302.5540us  Device To Host
      24         -         -         -           -  2.066576ms  Gpu page fault groups
Total CPU Page faults: 30
```

512 GSL w/ Prefetch

```
!nvprof ./imageLoad
```

```
==23637== NVPROF is profiling process 23637, command: ./imageLoad
Image loaded successfully!
Image dimensions: 512x512
Array size: 1048576
Will Process Image
Done Process Image
Image Saved
Validation from image
MSE: 2.3412e-05
RMSE from image: 0.00483859
Validation from file
MSE: 1.06483e-05
RMSE from file 0.00326317
==23637== Profiling application: ./imageLoad
==23637== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:  100.00%  77.206ms        10  7.7206ms  7.6546ms  8.2211ms  GaussianBlur(GaussianParams*)
      API calls:   55.84%  100.52ms         7  14.360ms  2.4640us  100.41ms  cudaMallocManaged
                   42.84%  77.108ms         1  77.108ms  77.108ms  77.108ms  cudaDeviceSynchronize
                    0.73%  1.3060ms         8  163.25us  60.170us  487.48us  cudaMemPrefetchAsync
                    0.30%  548.49us         7  78.355us  14.221us  113.37us  cudaFree
                    0.19%  349.79us        10  34.979us  5.2060us  236.64us  cudaLaunchKernel
                    0.08%  149.70us       114  1.3130us     102ns  66.655us  cuDeviceGetAttribute
                    0.01%  12.562us         1  12.562us  12.562us  12.562us  cuDeviceGetName
                    0.00%  6.1240us         1  6.1240us  6.1240us  6.1240us  cudaGetDevice
                    0.00%  4.8690us         1  4.8690us  4.8690us  4.8690us  cuDeviceGetPCIBusId
                    0.00%  1.6070us         3     535ns     133ns  1.1640us  cuDeviceGetCount
                    0.00%     843ns         2     421ns     182ns     661ns  cuDeviceGet
                    0.00%     564ns         1     564ns     564ns     564ns  cudaGetLastError
                    0.00%     517ns         1     517ns     517ns     517ns  cuDeviceTotalMem
                    0.00%     322ns         1     322ns     322ns     322ns  cuModuleGetLoadingMode
                    0.00%     278ns         1     278ns     278ns     278ns  cuDeviceGetUuid

==23637== Unified Memory profiling result:
Device "Tesla T4 (0)"
   Count  Avg Size  Min Size  Max Size  Total Size  Total Time  Name
      21  244.19KB  4.0000KB  1.0039MB  5.007813MB  482.6440us  Host To Device
       7  586.86KB  4.0000KB  1.0000MB  4.011719MB  330.6160us  Device To Host
      22         -         -         -           -  1.760440ms  Gpu page fault groups
Total CPU Page faults: 17
```

512 GSL w/ Prefetch & Page Creation

```
!nvprof ./imageLoad
```

```
==6459== NVPROF is profiling process 6459, command: ./imageLoad
Image loaded successfully!
Image dimensions: 512x512
Array size: 1048576
Will Process Image
Done Process Image
Image Saved
Validation from image
MSE: 2.3412e-05
RMSE from image: 0.00483859
Validation from file
MSE: 1.06483e-05
RMSE from file 0.00326317
==6459== Profiling application: ./imageLoad
==6459== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:  100.00%  76.603ms        10  7.6603ms  7.6464ms  7.6750ms  GaussianBlur(GaussianParams*)
      API calls:   62.22%  136.11ms         7  19.444ms  2.4650us  135.99ms  cudaMallocManaged
                   35.01%  76.580ms         1  76.580ms  76.580ms  76.580ms  cudaDeviceSynchronize
                    1.06%  2.3100ms        16  144.38us  3.3200us  1.0298ms  cudaMemPrefetchAsync
                    0.76%  1.6634ms         7  237.63us  15.645us  1.2151ms  cudaFree
                    0.45%  986.65us         1  986.65us  986.65us  986.65us  cuDeviceGetPCIBusId
                    0.41%  906.40us        10  90.639us  3.5980us  868.38us  cudaLaunchKernel
                    0.07%  158.53us       114  1.3900us     123ns  57.859us  cuDeviceGetAttribute
                    0.01%  19.533us         3  6.5110us     157ns  19.151us  cuDeviceGetCount
                    0.01%  13.124us         1  13.124us  13.124us  13.124us  cuDeviceTotalMem
                    0.01%  11.898us         1  11.898us  11.898us  11.898us  cuDeviceGetName
                    0.00%  6.2800us         1  6.2800us  6.2800us  6.2800us  cudaGetDevice
                    0.00%  1.3260us         2     663ns     168ns  1.1580us  cuDeviceGet
                    0.00%     585ns         1     585ns     585ns     585ns  cuModuleGetLoadingMode
                    0.00%     302ns         1     302ns     302ns     302ns  cudaGetLastError
                    0.00%     246ns         1     246ns     246ns     246ns  cuDeviceGetUuid

==6459== Unified Memory profiling result:
Device "Tesla T4 (0)"
   Count  Avg Size  Min Size  Max Size  Total Size  Total Time  Name
       5  616.00KB  4.0000KB  1.0000MB  3.007813MB  273.1130us  Host To Device
       5  616.00KB  4.0000KB  1.0000MB  3.007813MB  247.3850us  Device To Host
Total CPU Page faults: 16
```

512 GSL w/ Prefetch, Page Creation & Mem. Advise

```
!nvprof ./imageLoad
```

```
==8125== NVPROF is profiling process 8125, command: ./imageLoad
Image loaded successfully!
Image dimensions: 512x512
Array size: 1048576
Will Process Image
Done Process Image
Image Saved
Validation from image
MSE: 2.3412e-05
RMSE from image: 0.00483859
Validation from file
MSE: 1.06483e-05
RMSE from file 0.00326317
==8125== Profiling application: ./imageLoad
==8125== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:  100.00%  76.607ms        10  7.6607ms  7.6558ms  7.6680ms  GaussianBlur(GaussianParams*)
      API calls:   57.41%  107.74ms         7  15.392ms  2.4250us  107.62ms  cudaMallocManaged
                   40.81%  76.592ms         1  76.592ms  76.592ms  76.592ms  cudaDeviceSynchronize
                    0.70%  1.3136ms        16  82.098us  3.1080us  543.20us  cudaMemPrefetchAsync
                    0.45%  838.78us        10  83.878us  3.2930us  801.37us  cudaLaunchKernel
                    0.35%  649.88us         7  92.840us  14.455us  191.16us  cudaFree
                    0.17%  326.45us         8  40.806us  12.243us  70.910us  cudaMemAdvise
                    0.09%  169.98us       114  1.4910us     108ns  61.512us  cuDeviceGetAttribute
                    0.01%  14.119us         1  14.119us  14.119us  14.119us  cuDeviceGetName
                    0.00%  5.4720us         1  5.4720us  5.4720us  5.4720us  cuDeviceGetPCIBusId
                    0.00%  2.4500us         1  2.4500us  2.4500us  2.4500us  cudaGetDevice
                    0.00%  1.4860us         3     495ns     134ns  1.0810us  cuDeviceGetCount
                    0.00%     831ns         2     415ns     146ns     685ns  cuDeviceGet
                    0.00%     636ns         1     636ns     636ns     636ns  cuDeviceTotalMem
                    0.00%     538ns         1     538ns     538ns     538ns  cuModuleGetLoadingMode
                    0.00%     433ns         1     433ns     433ns     433ns  cudaGetLastError
                    0.00%     290ns         1     290ns     290ns     290ns  cuDeviceGetUuid

==8125== Unified Memory profiling result:
Device "Tesla T4 (0)"
   Count  Avg Size  Min Size  Max Size  Total Size  Total Time  Name
       4  769.00KB  4.0000KB  1.0000MB  3.003906MB  272.0260us  Host To Device
       3  1.0000MB  1.0000MB  1.0000MB  3.000000MB  243.8020us  Device To Host
Total CPU Page faults: 15
```

nvprof Data 1024x1024

1024 GSL

```
!nvprof ./imageLoad
```

```
==26027== NVPROF is profiling process 26027, command: ./imageLoad
Image loaded successfully!
Image dimensions: 1024x1024
Will Process Image
Finished processing.
1024x1024_outputImage.jpg saved.
Validation from image
MSE: 1.5554e-05
RMSE from image: 0.00394385
Validation from file
MSE: 4.27375e-06
RMSE from file 0.00206731
==26027== Profiling application: ./imageLoad
==26027== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:  100.00%  245.84ms        10  24.584ms  14.735ms  32.704ms  GaussianBlur(GaussianParams*)
      API calls:   70.13%  245.84ms         1  245.84ms  245.84ms  245.84ms  cudaDeviceSynchronize
                   29.42%  103.14ms         7  14.735ms  10.296us  103.03ms  cudaMallocManaged
                    0.34%  1.1788ms         7  168.40us  76.044us  255.89us  cudaFree
                    0.06%  211.97us        10  21.196us  3.4660us  174.49us  cudaLaunchKernel
                    0.04%  143.74us       114  1.2600us     104ns  60.822us  cuDeviceGetAttribute
                    0.00%  14.293us         1  14.293us  14.293us  14.293us  cuDeviceGetName
                    0.00%  4.8830us         1  4.8830us  4.8830us  4.8830us  cuDeviceGetPCIBusId
                    0.00%  1.2890us         3     429ns     122ns     901ns  cuDeviceGetCount
                    0.00%     814ns         2     407ns     165ns     649ns  cuDeviceGet
                    0.00%     517ns         1     517ns     517ns     517ns  cuDeviceTotalMem
                    0.00%     515ns         1     515ns     515ns     515ns  cuModuleGetLoadingMode
                    0.00%     448ns         1     448ns     448ns     448ns  cudaGetLastError
                    0.00%     256ns         1     256ns     256ns     256ns  cuDeviceGetUuid

==26027== Unified Memory profiling result:
Device "Tesla T4 (0)"
   Count  Avg Size  Min Size  Max Size  Total Size  Total Time  Name
      74  166.92KB  4.0000KB  0.9961MB  12.06250MB  1.239462ms  Host To Device
      74  166.92KB  4.0000KB  0.9961MB  12.06250MB  1.084800ms  Device To Host
      46         -         -         -           -  5.250723ms  Gpu page fault groups
Total CPU Page faults: 74
```

1024 GSL w/ Prefetch

```
!nvprof ./imageLoad
```

```
==23432== NVPROF is profiling process 23432, command: ./imageLoad
Image loaded successfully!
Image dimensions: 1024x1024
Array size: 4194304
Will Process Image
Done Process Image
Image Saved
Validation from image
MSE: 1.5554e-05
RMSE from image: 0.00394385
Validation from file
MSE: 4.27375e-06
RMSE from file 0.00206731
==23432== Profiling application: ./imageLoad
==23432== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:  100.00%  248.67ms        10  24.867ms  21.502ms  32.184ms  GaussianBlur(GaussianParams*)
      API calls:   70.31%  248.59ms         1  248.59ms  248.59ms  248.59ms  cudaDeviceSynchronize
                   27.86%  98.508ms         7  14.073ms  8.7820us  98.394ms  cudaMallocManaged
                    0.85%  3.0022ms        10  300.22us  5.3220us  2.8979ms  cudaLaunchKernel
                    0.66%  2.3187ms         8  289.84us  6.4600us  544.67us  cudaMemPrefetchAsync
                    0.27%  942.07us         7  134.58us  55.213us  201.18us  cudaFree
                    0.05%  178.88us       114  1.5690us     106ns  93.179us  cuDeviceGetAttribute
                    0.00%  12.233us         1  12.233us  12.233us  12.233us  cuDeviceGetName
                    0.00%  5.6870us         1  5.6870us  5.6870us  5.6870us  cuDeviceGetPCIBusId
                    0.00%  4.3200us         1  4.3200us  4.3200us  4.3200us  cudaGetDevice
                    0.00%  1.7080us         3     569ns     120ns  1.2680us  cuDeviceGetCount
                    0.00%     645ns         2     322ns     235ns     410ns  cuDeviceGet
                    0.00%     594ns         1     594ns     594ns     594ns  cuDeviceTotalMem
                    0.00%     549ns         1     549ns     549ns     549ns  cudaGetLastError
                    0.00%     449ns         1     449ns     449ns     449ns  cuModuleGetLoadingMode
                    0.00%     335ns         1     335ns     335ns     335ns  cuDeviceGetUuid

==23432== Unified Memory profiling result:
Device "Tesla T4 (0)"
   Count  Avg Size  Min Size  Max Size  Total Size  Total Time  Name
       9  1.3407MB  4.0000KB  2.0000MB  12.06641MB  1.062568ms  Host To Device
       8  1.5010MB  4.0000KB  2.0000MB  12.00781MB  967.4990us  Device To Host
      24         -         -         -           -  3.009018ms  Gpu page fault groups
Total CPU Page faults: 38
```

1024 GSL w/ Prefetch & Page Creation

```
!nvprof ./imageLoad

==6895== NVPROF is profiling process 6895, command: ./imageLoad
Image loaded successfully!
Image dimensions: 1024x1024
Array size: 4194304
Will Process Image
Done Process Image
Image Saved
Validation from image
MSE: 1.5554e-05
RMSE from image: 0.00394385
Validation from file
MSE: 4.27375e-06
RMSE from file 0.00206731
==6895== Profiling application: ./imageLoad
==6895== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:  100.00%  238.50ms        10  23.850ms  15.059ms  30.348ms  GaussianBlur(GaussianParams*)
      API calls:   67.57%  238.48ms         1  238.48ms  238.48ms  238.48ms  cudaDeviceSynchronize
                   31.13%  109.86ms         7  15.694ms  9.3190us  109.73ms  cudaMallocManaged
                    0.51%  1.8139ms        16  113.37us  3.1020us  461.07us  cudaMemPrefetchAsync
                    0.49%  1.7176ms        10  171.76us  3.4530us  1.6794ms  cudaLaunchKernel
                    0.25%  888.35us         7  126.91us  52.635us  220.61us  cudaFree
                    0.04%  153.69us       114  1.3480us     106ns  70.529us  cuDeviceGetAttribute
                    0.00%  13.189us         1  13.189us  13.189us  13.189us  cuDeviceGetName
                    0.00%  5.6820us         1  5.6820us  5.6820us  5.6820us  cuDeviceGetPCIBusId
                    0.00%  5.0290us         1  5.0290us  5.0290us  5.0290us  cudaGetDevice
                    0.00%  1.5840us         3     528ns     112ns  1.1710us  cuDeviceGetCount
                    0.00%     962ns         2     481ns     197ns     765ns  cuDeviceGet
                    0.00%     763ns         1     763ns     763ns     763ns  cuDeviceTotalMem
                    0.00%     599ns         1     599ns     599ns     599ns  cuModuleGetLoadingMode
                    0.00%     406ns         1     406ns     406ns     406ns  cudaGetLastError
                    0.00%     334ns         1     334ns     334ns     334ns  cuDeviceGetUuid

==6895== Unified Memory profiling result:
Device "Tesla T4 (0)"
   Count  Avg Size  Min Size  Max Size  Total Size  Total Time  Name
       8  1.5010MB  4.0000KB  2.0000MB  12.00781MB  1.058950ms  Host To Device
       8  1.5010MB  4.0000KB  2.0000MB  12.00781MB  965.3220us  Device To Host
Total CPU Page faults: 38
```

1024 GSL w/ Prefetch, Page Creation & Mem. Advise

```
!nvprof ./imageLoad
```

```
==8602== NVPROF is profiling process 8602, command: ./imageLoad
Image loaded successfully!
Image dimensions: 1024x1024
Array size: 4194304
Will Process Image
Done Process Image
Image Saved
Validation from image
MSE: 1.5554e-05
RMSE from image: 0.00394385
Validation from file
MSE: 4.27375e-06
RMSE from file 0.00206731
==8602== Profiling application: ./imageLoad
==8602== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:  100.00%  235.24ms        10  23.524ms  15.063ms  30.329ms  GaussianBlur(GaussianParams*)
      API calls:   62.48%  235.20ms         1  235.20ms  235.20ms  235.20ms  cudaDeviceSynchronize
                   35.68%  134.31ms         7  19.187ms  13.120us  134.14ms  cudaMallocManaged
                    0.63%  2.3732ms        16  148.32us  3.9750us  560.14us  cudaMemPrefetchAsync
                    0.42%  1.5952ms         7  227.89us  170.87us  340.79us  cudaFree
                    0.38%  1.4451ms         8  180.63us  17.576us  406.77us  cudaMemAdvise
                    0.36%  1.3364ms        10  133.64us  5.4740us  1.2773ms  cudaLaunchKernel
                    0.04%  147.18us       114  1.2910us     106ns  57.741us  cuDeviceGetAttribute
                    0.00%  14.517us         1  14.517us  14.517us  14.517us  cuDeviceGetName
                    0.00%  7.5870us         1  7.5870us  7.5870us  7.5870us  cuDeviceGetPCIBusId
                    0.00%  6.2540us         1  6.2540us  6.2540us  6.2540us  cudaGetDevice
                    0.00%  1.3380us         3     446ns     134ns     935ns  cuDeviceGetCount
                    0.00%  1.1060us         2     553ns     230ns     876ns  cuDeviceGet
                    0.00%     575ns         1     575ns     575ns     575ns  cuModuleGetLoadingMode
                    0.00%     545ns         1     545ns     545ns     545ns  cudaGetLastError
                    0.00%     416ns         1     416ns     416ns     416ns  cuDeviceTotalMem
                    0.00%     299ns         1     299ns     299ns     299ns  cuDeviceGetUuid

==8602== Unified Memory profiling result:
Device "Tesla T4 (0)"
   Count  Avg Size  Min Size  Max Size  Total Size  Total Time  Name
       7  1.7148MB  4.0000KB  2.0000MB  12.00391MB  1.049413ms  Host To Device
       6  2.0000MB  2.0000MB  2.0000MB  12.00000MB  962.2810us  Device To Host
Total CPU Page faults: 37
```

nvprof Data 2048x2048

2048 GSL

```
!nvprof ./imageLoad
```

```
==23876== NVPROF is profiling process 23876, command: ./imageLoad
Image loaded successfully!
Image dimensions: 2048x2048
Will Process Image
Finished processing.
2048x2048_outputImage_cuda_a.jpg saved.
Validation from image
MSE: 1.35444e-05
RMSE from image: 0.00368026
Validation from file
MSE: 2.09942e-06
RMSE from file 0.00144894
==23876== Profiling application: ./imageLoad
==23876== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:  100.00%  711.45ms        10  71.145ms  60.339ms  131.66ms  GaussianBlur(GaussianParams*)
      API calls:   85.93%  711.41ms         1  711.41ms  711.41ms  711.41ms  cudaDeviceSynchronize
                   13.34%  110.42ms         7  15.774ms  8.0620us  110.25ms  cudaMallocManaged
                    0.68%  5.6263ms         7  803.76us  91.580us  1.0365ms  cudaFree
                    0.03%  259.95us        10  25.994us  5.3040us  203.56us  cudaLaunchKernel
                    0.02%  145.22us       114  1.2730us     106ns  64.089us  cuDeviceGetAttribute
                    0.00%  14.255us         1  14.255us  14.255us  14.255us  cuDeviceGetName
                    0.00%  5.1980us         1  5.1980us  5.1980us  5.1980us  cuDeviceGetPCIBusId
                    0.00%  1.6700us         3     556ns     208ns  1.1870us  cuDeviceGetCount
                    0.00%     848ns         2     424ns     212ns     636ns  cuDeviceGet
                    0.00%     699ns         1     699ns     699ns     699ns  cuDeviceTotalMem
                    0.00%     513ns         1     513ns     513ns     513ns  cudaGetLastError
                    0.00%     348ns         1     348ns     348ns     348ns  cuModuleGetLoadingMode
                    0.00%     281ns         1     281ns     281ns     281ns  cuDeviceGetUuid

==23876== Unified Memory profiling result:
Device "Tesla T4 (0)"
   Count  Avg Size  Min Size  Max Size  Total Size  Total Time  Name
     299  164.60KB  4.0000KB  0.9961MB  48.06250MB  4.972040ms  Host To Device
     290  169.71KB  4.0000KB  0.9961MB  48.06250MB  4.291386ms  Device To Host
     183         -         -         -           -  21.00900ms  Gpu page fault groups
Total CPU Page faults: 290
```

2048 GSL w/ Prefetch

```
!nvprof ./imageLoad
```

```
==23213== NVPROF is profiling process 23213, command: ./imageLoad
Image loaded successfully!
Image dimensions: 2048x2048
Array size: 16777216
Will Process Image
Done Process Image
Image Saved
Validation from image
MSE: 1.35444e-05
RMSE from image: 0.00368026
Validation from file
MSE: 2.09942e-06
RMSE from file 0.00144894
==23213== Profiling application: ./imageLoad
==23213== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:  100.00%  715.01ms        10  71.501ms  61.791ms  124.85ms  GaussianBlur(GaussianParams*)
      API calls:   86.06%  714.97ms         1  714.97ms  714.97ms  714.97ms  cudaDeviceSynchronize
                   12.16%  101.05ms         7  14.436ms  8.7370us  100.92ms  cudaMallocManaged
                    0.64%  5.3091ms         8  663.63us  7.1050us  1.4764ms  cudaMemPrefetchAsync
                    0.62%  5.1540ms         7  736.28us  135.00us  2.2917ms  cudaFree
                    0.50%  4.1673ms        10  416.73us  5.5630us  4.1098ms  cudaLaunchKernel
                    0.02%  139.88us       114  1.2260us     105ns  53.161us  cuDeviceGetAttribute
                    0.00%  13.198us         1  13.198us  13.198us  13.198us  cuDeviceGetName
                    0.00%  5.7770us         1  5.7770us  5.7770us  5.7770us  cuDeviceGetPCIBusId
                    0.00%  5.1130us         1  5.1130us  5.1130us  5.1130us  cudaGetDevice
                    0.00%  1.2930us         3     431ns     119ns     828ns  cuDeviceGetCount
                    0.00%  1.1290us         2     564ns     240ns     889ns  cuDeviceGet
                    0.00%     629ns         1     629ns     629ns     629ns  cuDeviceTotalMem
                    0.00%     422ns         1     422ns     422ns     422ns  cudaGetLastError
                    0.00%     394ns         1     394ns     394ns     394ns  cuModuleGetLoadingMode
                    0.00%     263ns         1     263ns     263ns     263ns  cuDeviceGetUuid

==23213== Unified Memory profiling result:
Device "Tesla T4 (0)"
   Count  Avg Size  Min Size  Max Size  Total Size  Total Time  Name
      27  1.7802MB  4.0000KB  2.0000MB  48.06641MB  4.193978ms  Host To Device
      26  1.8465MB  4.0000KB  2.0000MB  48.00781MB  3.852323ms  Device To Host
      96         -         -         -           -  7.435883ms  Gpu page fault groups
Total CPU Page faults: 146
```

2048 GSL w/ Prefetch & Page Creation

```
!nvprof ./imageLoad

==7234== NVPROF is profiling process 7234, command: ./imageLoad
Image loaded successfully!
Image dimensions: 2048x2048
Array size: 16777216
Will Process Image
Done Process Image
Image Saved
Validation from image
MSE: 1.35444e-05
RMSE from image: 0.00368026
Validation from file
MSE: 2.09942e-06
RMSE from file 0.00144894
==7234== Profiling application: ./imageLoad
==7234== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:  100.00%  702.13ms        10  70.213ms  60.632ms  121.42ms  GaussianBlur(GaussianParams*)
      API calls:   82.23%  702.01ms         1  702.01ms  702.01ms  702.01ms  cudaDeviceSynchronize
                   15.77%  134.65ms         7  19.236ms  11.059us  134.44ms  cudaMallocManaged
                    0.75%  6.4070ms        16  400.44us  3.9640us  1.7541ms  cudaMemPrefetchAsync
                    0.74%  6.2892ms        10  628.92us  5.2030us  6.1534ms  cudaLaunchKernel
                    0.49%  4.1806ms         7  597.23us  163.90us  1.9132ms  cudaFree
                    0.02%  178.65us       114  1.5670us     148ns  69.705us  cuDeviceGetAttribute
                    0.00%  15.329us         1  15.329us  15.329us  15.329us  cuDeviceGetName
                    0.00%  6.9400us         1  6.9400us  6.9400us  6.9400us  cuDeviceGetPCIBusId
                    0.00%  5.9070us         1  5.9070us  5.9070us  5.9070us  cudaGetDevice
                    0.00%  1.9920us         3     664ns     233ns  1.4850us  cuDeviceGetCount
                    0.00%     790ns         2     395ns     220ns     570ns  cuDeviceGet
                    0.00%     710ns         1     710ns     710ns     710ns  cuDeviceTotalMem
                    0.00%     489ns         1     489ns     489ns     489ns  cudaGetLastError
                    0.00%     446ns         1     446ns     446ns     446ns  cuModuleGetLoadingMode
                    0.00%     354ns         1     354ns     354ns     354ns  cuDeviceGetUuid

==7234== Unified Memory profiling result:
Device "Tesla T4 (0)"
   Count  Avg Size  Min Size  Max Size  Total Size  Total Time  Name
      28  1.7168MB  4.0000KB  2.0000MB  48.07031MB  4.193435ms  Host To Device
      27  1.7782MB  4.0000KB  2.0000MB  48.01172MB  3.854532ms  Device To Host
       1         -         -         -           -  81.43800us  Gpu page fault groups
Total CPU Page faults: 147
```

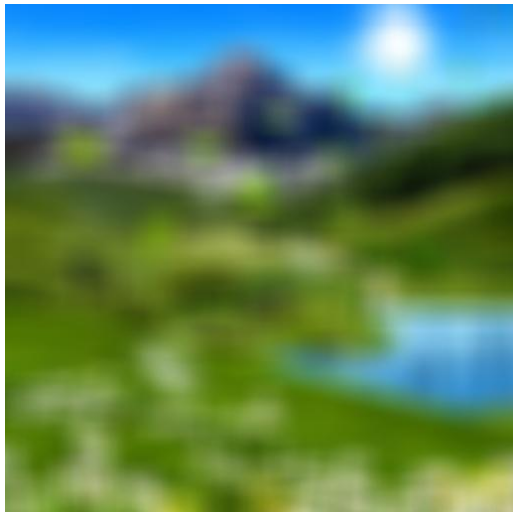
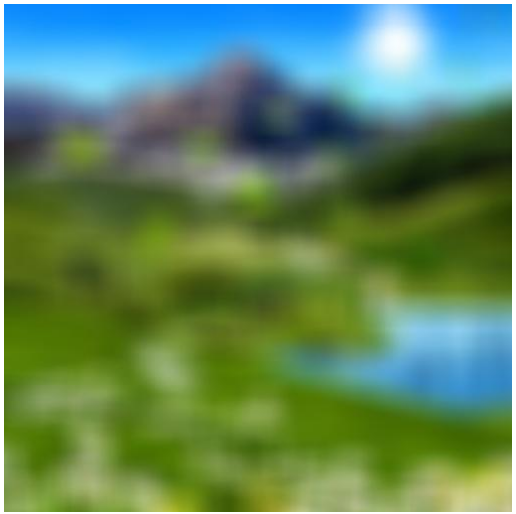
2048 GSL w/ Prefetch, Page Creation & Mem. Advise

```
!nvprof ./imageLoad
```

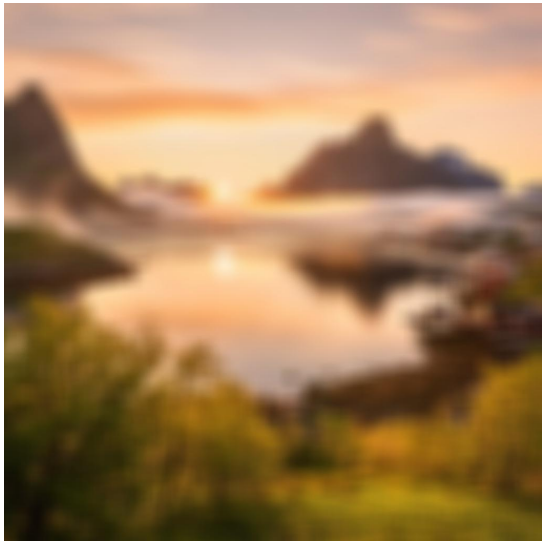
```
==8991== NVPROF is profiling process 8991, command: ./imageLoad
Image loaded successfully!
Image dimensions: 2048x2048
Array size: 16777216
Will Process Image
Done Process Image
Image Saved
Validation from image
MSE: 1.35444e-05
RMSE from image: 0.00368026
Validation from file
MSE: 2.09942e-06
RMSE from file 0.00144894
==8991== Profiling application: ./imageLoad
==8991== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:  100.00%  716.32ms        10  71.632ms  60.948ms  121.34ms  GaussianBlur(GaussianParams*)
      API calls:   84.10%  716.30ms         1  716.30ms  716.30ms  716.30ms  cudaDeviceSynchronize
                   13.86%  118.09ms         7  16.870ms  9.0630us  117.92ms  cudaMallocManaged
                    0.57%  4.8603ms        10  486.03us  3.5270us  4.8212ms  cudaLaunchKernel
                    0.57%  4.8527ms        16  303.29us  3.3760us  1.4720ms  cudaMemPrefetchAsync
                    0.45%  3.8386ms         7  548.37us  385.62us  719.72us  cudaFree
                    0.43%  3.6225ms         8  452.81us  20.662us  1.0548ms  cudaMemAdvise
                    0.02%  143.38us       114  1.2570us     107ns  56.443us  cuDeviceGetAttribute
                    0.00%  12.808us         1  12.808us  12.808us  12.808us  cuDeviceGetName
                    0.00%  7.2780us         1  7.2780us  7.2780us  7.2780us  cudaGetDevice
                    0.00%  5.5140us         1  5.5140us  5.5140us  5.5140us  cuDeviceGetPCIBusId
                    0.00%  1.3390us         3     446ns     146ns     881ns  cuDeviceGetCount
                    0.00%     821ns         2     410ns     186ns     635ns  cuDeviceGet
                    0.00%     671ns         1     671ns     671ns     671ns  cuDeviceTotalMem
                    0.00%     458ns         1     458ns     458ns     458ns  cudaGetLastError
                    0.00%     332ns         1     332ns     332ns     332ns  cuModuleGetLoadingMode
                    0.00%     279ns         1     279ns     279ns     279ns  cuDeviceGetUuid

==8991== Unified Memory profiling result:
Device "Tesla T4 (0)"
   Count  Avg Size  Min Size  Max Size  Total Size  Total Time  Name
      25  1.9202MB  4.0000KB  2.0000MB  48.00391MB  4.182204ms  Host To Device
      24  2.0000MB  2.0000MB  2.0000MB  48.00000MB  3.848360ms  Device To Host
Total CPU Page faults: 145
```

Output

512 x 512

Source	C Kernel	CUDA
		

1024 x 1024

Source	C Kernel	CUDA
		

2048 x 2048

Source	C Kernel	CUDA