Replacing deconv to conv: pre_replace_deconv pass

Felix Kim

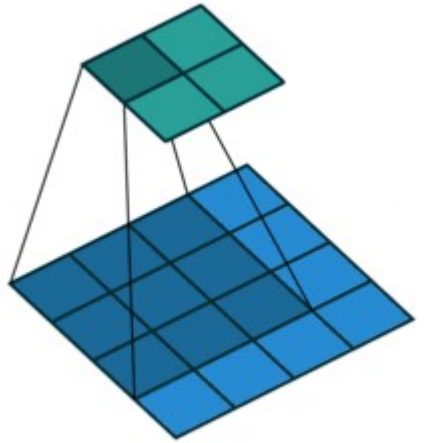
Intro

- Some deconv testcases failed in accuracy with onednn
  - Surprisingly, no deconvolution implementations were used!
  - After some debugging, I found the little known pass: pre_replace_deconv
- Deconvolution can be replaced to convolution in some cases
  - How?
  - Why?

Operations Review

Convolution

- Meaning
  - One of downsampling method
  - Compress the input data into an abstract spatial representation
- Usecase
  - Almost all network

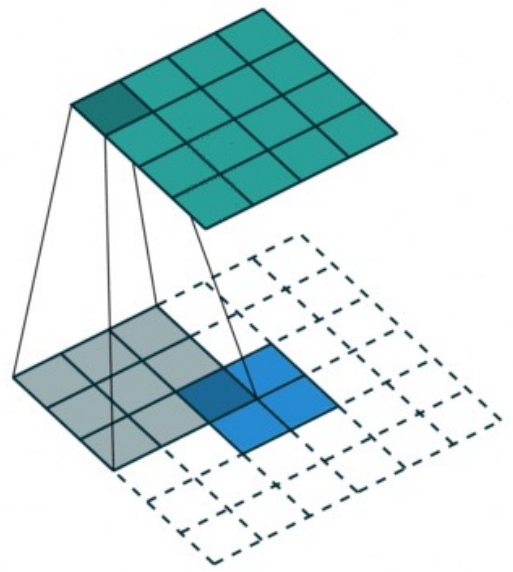


Deconvolution

- Meaning
    - Revert convolution
- Usecase
    - Signal Processing

Transposed Convolution

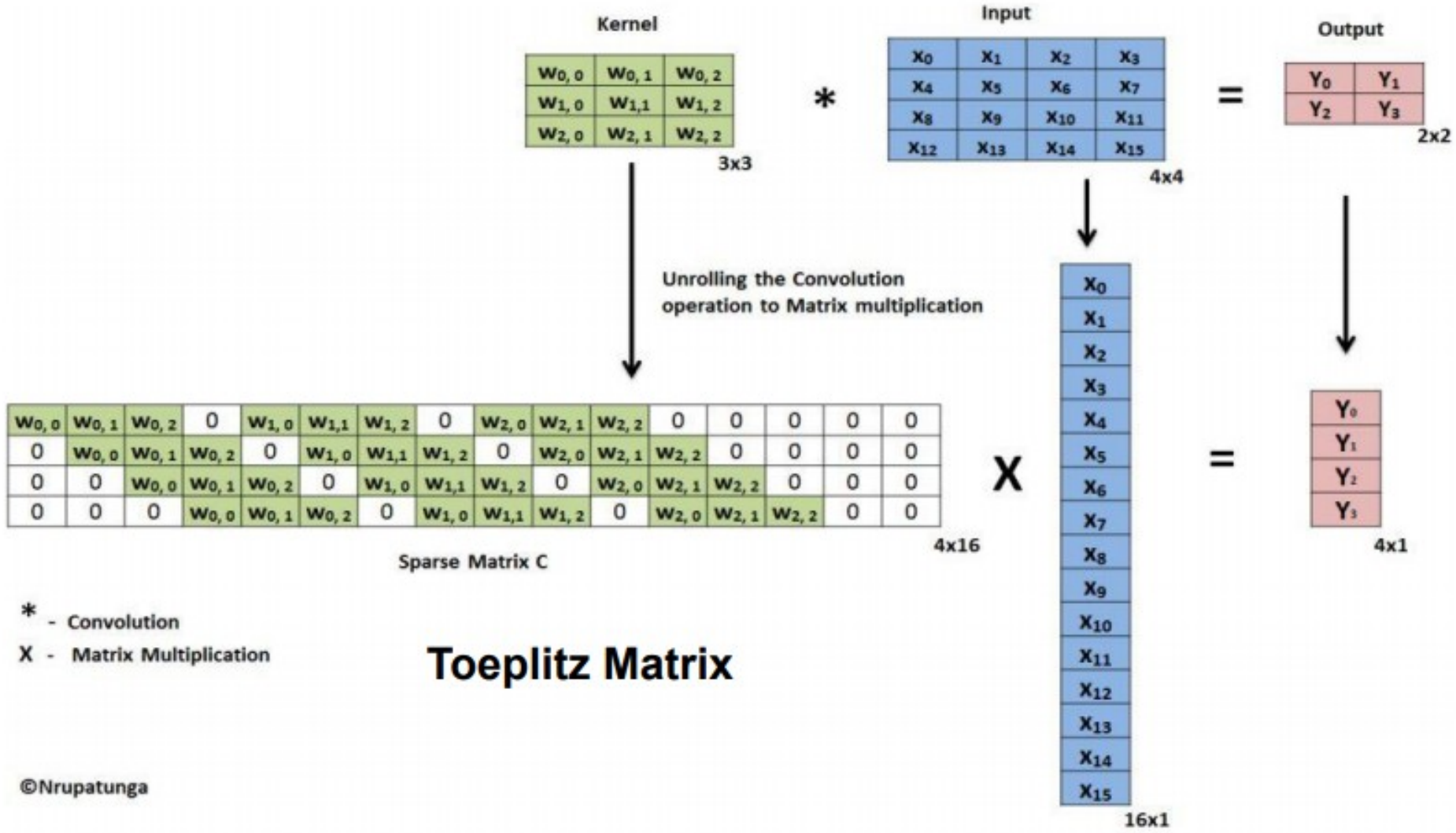
- Meaning
  - One of upsampling techniques
  - Decompress the abstract representation into something of use.
  - Deconvolution with unknown weight which has to be learned
- Usecase
  - Super Resolution, Semantic Segmentation, …



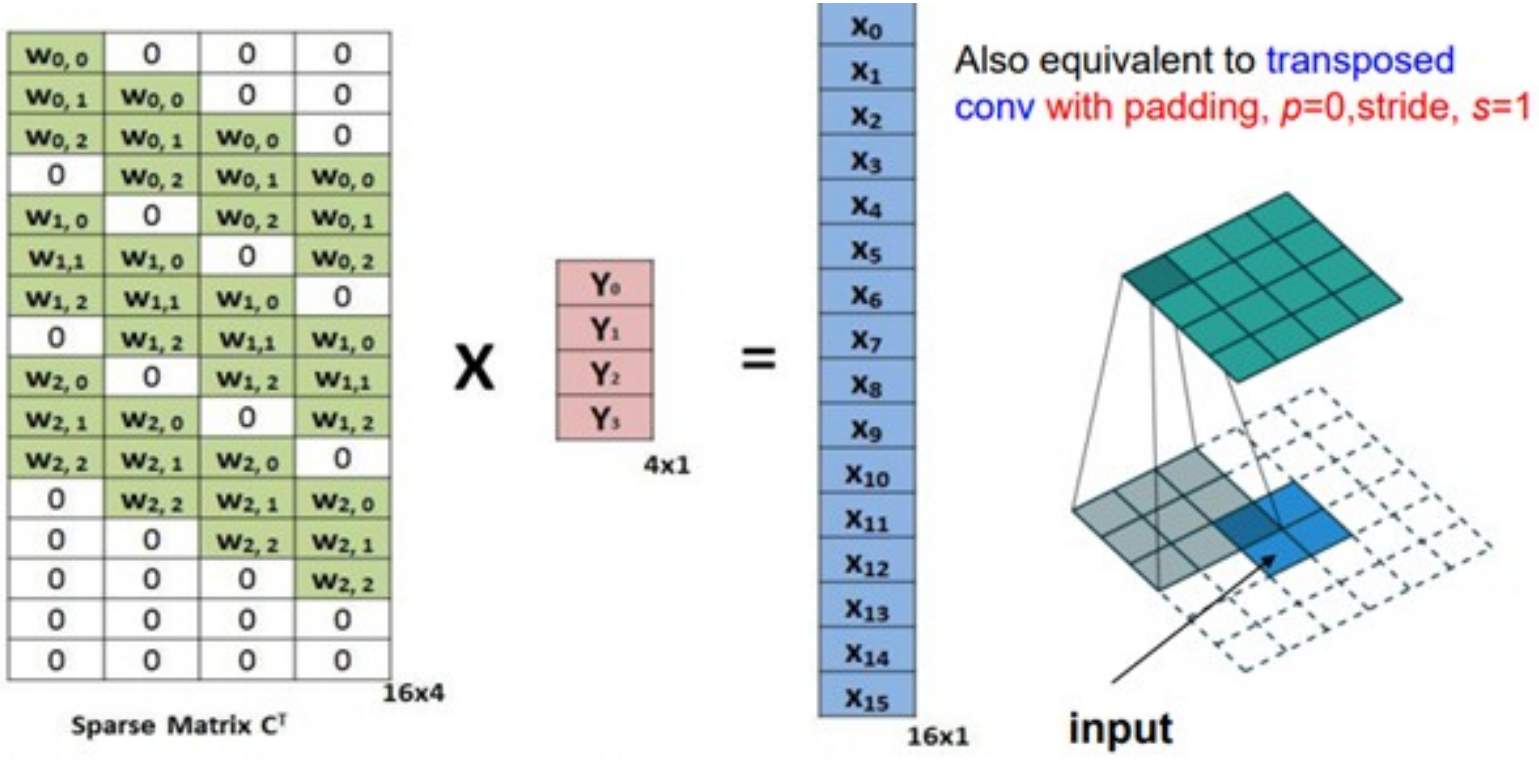
Deconvolution vs Transposed Convolution

- Common
    - Same shape(Input/Kernel/Output)
- Difference
    - Deconvolution
        - exact inverse operation of convolution
    - Transposed convolution
        - Can be trained to behave inverse of convolution, but not necessarily.
-
-

Why it called “Transposed” Convolution



Why it called “Transposed” Convolution



Notes:

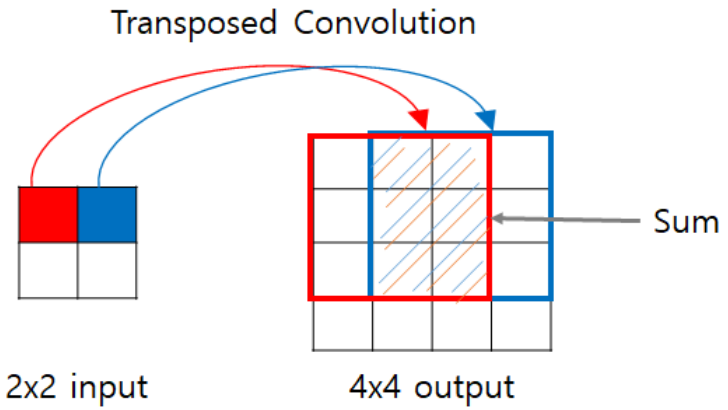
* Full rank matrix always have (right or left) generalized inverse matrix
* Value of matrix has no meaning in transposed convolution.

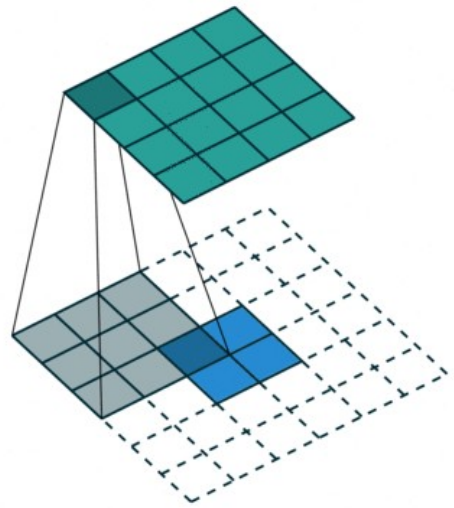
Deconv to Conv

How replace deconv to conv?

- `limit optimization to stride = 1`
- Add paddings
- rotate weight 180

Two ways to calculate deconv





a	b	11	12	13	a11	a12+b11	a13+b12	b13
c	d	21	22	23	a21	a22+b21	a23+b22	b23
		31	32	33	a31	a32+b31	a33+b32	b33
					0	0	0	0

a33	a32	a31	0
a23	a22	a21	0
a13	a12	a11	0
0	0	0	0

0	b33	b32	b31
0	b23	b22	b21
0	b13	b12	b11
0	0	0	0

Related codes

pre_replace_deconv.cpp

```
for (size_t i = 0; i < spatial_rank; i++) {
    pad[i] = (filter_layout.spatial(spatial_rank - i - 1) - 1) - std::abs(pad[i]);
}       Vladimir Paramuzov, 7 months ago • [GPU] Update data structures for conv/poo
```

```
program_node& new_node = p.get_or_create(conv_prim);

auto& conv_node = new_node.as<convolution>();
conv_node.set_transposed(true);       Andrei Molotkov,
```

convolution_onednn.cpp

```cpp
static primitive_impl* create(const convolution_node& arg, const kernel_impl_params& impl_params) {
    auto& engine = impl_params.prog.get_engine();
    auto desc = get_convolution_descriptor(impl_params);
    auto attr = get_primitive_attributes(arg);
    dnnl::primitive_desc prim_desc{&desc->data, attr.get(), engine.get_onednn_engine(), nullptr};

    return new convolution_onednn(engine, desc, attr, prim_desc, get_weights_reorder(impl_params, prim_desc, arg.get_transposed()));
}
```

```cpp
static kernel_selector::WeightsReorderParams get_weights_reorder(const kernel_impl_params& impl_params, const dnnl::primitive_desc& pd, bool rotate) {
    kernel_selector::WeightsReorderParams weights_reorder_params;
    auto& reorderKS = kernel_selector::ReorderWeightsKernelSelctor::Instance();
    kernel_selector::reorder_weights_params r_params;

    //생략

    r_params.rotate_180 = rotate;       You, yesterday • Bugfix: deconv 2 conv optimization …

    //생략

    weights_reorder_params.engine = kernel_selector::WeightsReorderParams::Engine::GPU;
    weights_reorder_params.clKernel = std::make_shared<kernel_selector::clKernelData>(kernels_data[0].kernels[0]);
    weights_reorder_params.dest = r_params.output;

    return weights_reorder_params;
}
```

reorder_weights.cl

```
#if !REORDER_ROTATE      Alexey Suhov, 2 years ago • publish master branch snapshot, revision 49482ae3…
    uint output_idx = FUNC_CALL(get_output_index)(g, o, i, z, y, x);
#else
    uint output_idx = FUNC_CALL(get_output_index)(g, o, i, OUTPUT_SIZE_Z - z - 1, OUTPUT_SIZE_Y - y - 1, OUTPUT_SIZE_X - x - 1);
#endif
```

References

- https://medium.com/@marsxiang/convolutions-transposed-and-deconvolution-6430c358a5b6
- https://medium.com/apache-mxnet/transposed-convolutions-explained-with-ms-excel-52d13030c7e8
- https://naokishibuya.medium.com/up-sampling-with-transposed-convolution-9ae4f2df52d0
- https://towardsdatascience.com/understand-transposed-convolutions-and-build-your-own-transposed-convolution-layer-from-scratch-4f5d97b2967
- https://realblack0.github.io/2020/05/11/transpose-convolution.html
- https://analysisbugs.tistory.com/m/104