Auto-scheduler for TVM

Lianmin Zheng

Shanghai Jiao Tong University

Tensor Expression

$$C_{ij} = \sum_{k} A_{ik} B_{kj}$$

Tensor Expression

$$C_{ij} = \sum_{k} A_{ik} B_{kj}$$

AutoTVM Schedule



Tensor Expression

$$C_{ij} = \sum_{k} A_{ik} B_{kj}$$









Optimizer

Tensor Expression

$$C_{ij} = \sum_{k} A_{ik} B_{kj}$$









Optimizer





Tensor Expression

$$C_{ij} = \sum_{k} A_{ik} B_{kj}$$



AutoTVM Schedule













$$C_{ij} = \sum_{k} A_{ik} B_{kj}$$



AutoTVM Schedule



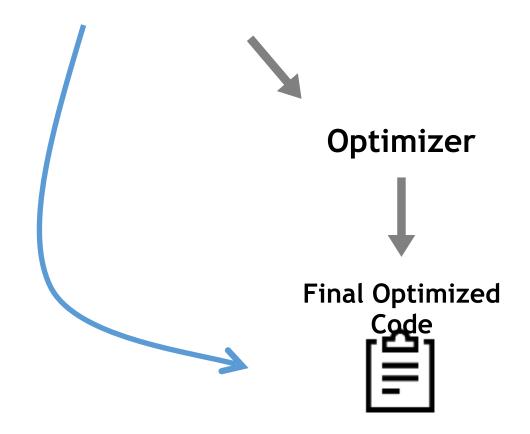






Tensor Expression

$$C_{ij} = \sum_{k} A_{ik} B_{kj}$$



Tensor Expression

$$C_{ij} = \sum_{k} A_{ik} B_{kj}$$







Optimizer





Tensor Expression

$$C_{N,CO,H,W} = \sum_{CI, KH, KW} A_{N,CO,H+KH, W+KW} B_{CO, CI,KH,KW}$$

Tensor Expression

$$C_{N,CO,H,W} = \sum_{CI, KH, KW} A_{N,CO,H+KH, W+KW} B_{CO, CI,KH,KW}$$

Auto-packing

- for vectorization
- for tensorization
- other packing strategies

Tensor Expression

$$C_{N,CO,H,W} = \sum_{CI, KH, KW} A_{N,CO,H+KH, W+KW} B_{CO, CI,KH,KW}$$

Auto-packing

- for vectorization
- for tensorization
- other packing strategies

Tile-knob on axes

Tune multi-level tiling on N, H, W, CO, CI

Tensor Expression

$$C_{N,CO,H,W} = \sum_{CI, KH, KW} A_{N,CO,H+KH, W+KW} B_{CO, CI,KH,KW}$$

Auto-packing

- for vectorization
- for tensorization
- other packing strategies

Tile-knob on axes

Tune multi-level tiling on N, H, W, CO, CI

Location-knob on buffers

Tune *compute_at* location for A, B, C and packing buffers

Tensor Expression

$$C_{N,CO,H,W} = \sum_{CI, KH, KW} A_{N,CO,H+KH, W+KW} B_{CO, CI,KH,KW}$$

Auto-packing

- for vectorization
- for tensorization
- other packing strategies

Tile-knob on axes

Tune multi-level tiling on N, H, W, CO, CI

Location-knob on buffers

Tune compute_at location for A, B, C and packing buffers

Annotation-knob on axes

Tune parallel, unroll

Tensor Expression

$$C_{N,CO,H,W} = \sum_{CI, KH, KW} A_{N,CO,H+KH, W+KW} B_{CO, CI,KH,KW}$$

Auto-packing

- for vectorization
- for tensorization
- other packing strategies

Tile-knob on axes

Tune multi-level tiling on N, H, W, CO, CI

Location-knob on buffers

Tune *compute_at* location for A, B, C and packing buffers

Annotation-knob on axes

Tune parallel, unroll

Map to different backends

Map threading, memory scoping, tensorization

Usage

tvm.create_schedule(C.op)

```
\rightarrow
```

autotvm.create schedule(C.op)

```
and contributions to be the fit, it is stold publication.

   Appropriation of the property of the page 1981.

        mount N = 1, "Only mention interpolars - 2 in this hopists".

   (ndpd) Mak(q, the House, site("Mad(ht y"))).

   a(n/p/f).stat(n, for.terns(_min("dashbina")))

             Artic - Nov. (Control Control 

   i(n/p/f).idat(ef, fee-termi_min("/fee-mi"))

             have I = \text{torophist}(\text{cond}(0), \Omega_1, \Omega_2, \Omega_3, \text{cond}(0))
            to a factor and the factor of the section and the distinct of their facility
                                                                                                                                                                                                                                                                                                                         D. s(m/p/d(stat)(q), the termi_min("decor"))
                                                                                                                                                                                                                                                                                                                                                      challed and the territorial decoration ("decoration)"

   specifically block (s), then the entire temperate ["The entire sur"]);

                                                                                                                                                                                                                                                                                                                                                        Parketh Making the Dead said to setting "to
          DVD-01 St. LOBY BIDLY (BCD4), CCD41
                                                                                                                                                                                                                                                                                                                       35 shortest Authorities, then the and noted the medical of the
             many space are control tagger many
                                                                                                                                                                                                                                                                                                                     30 S. GUTGUTEL FROMBRICK, ST., St., Co., VT., W., VC., TT., CA., CO., TJ., ML., Co.

   sprij overstjogsprant, mj.

            or, ey, or a signal) to relate print
                                                                                                                                                                                                                                                                                                                    Service of the servic
            of autobased archief.

a, d, g, a = a(0) spends.

             etgantinoppin (1936, Fr. 4), requirem ()
                                                                                                                                                                                                                                                                                                        D. Park Park (1997) apprendict parts
            (استعمل در راز والله والتوريط المدوات
                                                                                                                                                                                                                                                                                                            III out, now, rot orig("Mile_rot()apply(e, it, rot)
            Charles with Citie C. a. on adapted.
                                                                                                                                                                                                                                                                                                             III you got you styl "Miles to Jumping to the real"
          Countille estimation of the communication
                                                                                                                                                                                                                                                                                                                                                      cos_{ij} cos_{ij} cos_{ij} = cos_{ij} ("Life_{ij} y") cos_{ij} cos_{ij} (e_{ij} cos_{ij} cos_{ij})
          cheatile with Citie of the ran admired.
                                                                                                                                                                                                                                                                                                                                                       a 100 former des frances productions, more transported a real condition of the real conditions.
             ong neminal produce and produce the later than the state of the state 
            organizacjana, recent jegoven († 18. stp.
                                                                                                                                                                                                                                                                                                                       SHEET SHEET CONTRACTS AT CONTRACT, TOO I

   April 2000 Publishing Co.

                                                                                                                                                                                                                                                                                                                       and a specific community of the second
             projection of constant projection (2)
            containt, reprintative)
                                                                                                                                                                                                                                                                                                                     22 # corporables fatching
            data, magasta - pagasta, data
                                                                                                                                                                                                                                                                                                                       70 Air Smill to (Al., 160):
                                                                                                                                                                                                                                                                                                                                                                   \sigma_{\alpha} \cdot \Phi_{\alpha} \cdot \varphi_{\alpha} \cdot \alpha = \alpha (\lambda \cos \theta), appendix
                                                                                                                                                                                                                                                                                                                                                                            \operatorname{Autol} = \operatorname{Cons}(\operatorname{Autol}(n, A_{n,2n}, n))
                                                                                                                                                                                                                                                                                                                                                                           its, from a gloral aptitional, we have (-111a_{\mu}^{-1}) at a(1).
          E CORPOR TROOP, TORON.
                                                                                                                                                                                                                                                                                                                                                                            to Panel - attendication bands on two left like of trainers.
             ALC: A CHESS, DANGERTS, MINOR J. $150.
                                                                                                                                                                                                                                                                                                                                                                            to Foreign Month with the discovered Title of Labertin
            William technique (factor), "thereo", [T.D.
                                                                                                                                                                                                                                                                                                                                                                            afficially bindful, they the end adds "the endich of "the
          A. Alemany and Mr. (Seeding (003))
                                                                                                                                                                                                                                                                                                                                                                        STACKED LIGHTERY, CATACHORD EXCEL CONSIDERAL STA
        the management of the collection (002)
                                                                                                                                                                                                                                                                                                                                                                           ALLERS BOREN, THE PROPERTY PROPERTY IS NO
         v_{i},\, f_{i},\, g_{i},\, u = a) which is the same i
          bit, all, the discontinuous orbits, private to
                                                                                                                                                                                                                                                                                                                                                              s(ndpd), suppo(need, neps, folio, end (proping), etg)(nde, end (proping), end)\\
            to, or, is, of a staffilliant topologic solucity of

   a)min/d), angrafiared_maps, "seedd_mphinid", afg("seedd_mphinid") and);

          have a track and a platfilling of the coloring periods, where
          CANCELL STORM AND ADDRESS OF THE STORM OF TH
                                                                                                                                                                                                                                                                                                                                                      return of (regulate, larve), and (
```

```
def conv2d_no_batching(N, H, W, CO, CI, KH, KW, stride, padding):
   data = tvm.placeholder((N, CI, H, W), name='data')
   kennel = tvm.placeholder((CO, CI, KH, KW), name='kennel')
   conv = topi.nn.conv2d_nchw(data, kernel, stride, padding, dilation 1, out_dtype 'floatJ2')
   k = autolvm.create schedule([conv.op])
   return s, [raw_data, kernel, conv]
```

Goal and progress

Replace manual schedules in current TOPI without regression

Provide schedules for new operators: training ops, conv3d,

One month