

# Concatenation 层

- 初始示例代码
- axis

## 初始示例代码

```
import numpy as np
from cuda import cudart
import tensorrt as trt

nIn, cIn, hIn, wIn = 1, 3, 4, 5 # 输入张量 NCHW
data = np.arange(1, 1 + nIn * cIn * hIn * wIn, dtype=np.float32).reshape(nIn, cIn, hIn, wIn) # 输入数据

np.set_printoptions(precision=8, linewidth=200, suppress=True)
cudart.cudaDeviceSynchronize()

logger = trt.Logger(trt.Logger.ERROR)
builder = trt.Builder(logger)
network = builder.create_network(1 << int(trt.NetworkDefinitionCreationFlag.EXPLICIT_BATCH))
config = builder.create_builder_config()
inputT0 = network.add_input('inputT0', trt.DataType.FLOAT, (nIn, cIn, hIn, wIn))
#-----# 替换部分
concatenationLayer = network.add_concatenation([inputT0, inputT0])
#-----# 替换部分
network.mark_output(concatenationLayer.get_output(0))
engineString = builder.build_serialized_network(network, config)
engine = trt.Runtime(logger).deserialize_cuda_engine(engineString)
context = engine.create_execution_context()
_, stream = cudart.cudaStreamCreate()

inputH0 = np.ascontiguousarray(data.reshape(-1))
outputH0 = np.empty(context.get_binding_shape(1), dtype=trt.nptype(engine.get_binding_dtype(1)))
_, inputD0 = cudart.cudaMallocAsync(inputH0.nbytes, stream)
_, outputD0 = cudart.cudaMallocAsync(outputH0.nbytes, stream)

cudart.cudaMemcpyAsync(inputD0, inputH0.ctypes.data, inputH0.nbytes,
cudart.cudaMemcpyKind.cudaMemcpyHostToDevice, stream)
context.execute_async_v2([int(inputD0), int(outputD0)], stream)
cudart.cudaMemcpyAsync(outputH0.ctypes.data, outputD0, outputH0.nbytes,
cudart.cudaMemcpyKind.cudaMemcpyDeviceToHost, stream)
cudart.cudaStreamSynchronize(stream)

print("inputH0 :", data.shape)
print(data)
print("outputH0:", outputH0.shape)
print(outputH0)

cudart.cudaStreamDestroy(stream)
cudart.cudaFree(inputD0)
cudart.cudaFree(outputD0)
```

- 输入张量形状 (1,3,4,5)

$$\left[ \left[ \begin{bmatrix} 0. & 1. & 2. & 3. & 4. \\ 5. & 6. & 7. & 8. & 9. \\ 10. & 11. & 12. & 13. & 14. \\ 15. & 16. & 17. & 18. & 19. \end{bmatrix} \begin{bmatrix} 20. & 21. & 22. & 23. & 24. \\ 25. & 26. & 27. & 28. & 29. \\ 30. & 31. & 32. & 33. & 34. \\ 35. & 36. & 37. & 38. & 39. \end{bmatrix} \begin{bmatrix} 40. & 41. & 42. & 43. & 44. \\ 45. & 46. & 47. & 48. & 49. \\ 50. & 51. & 52. & 53. & 54. \\ 55. & 56. & 57. & 58. & 59. \end{bmatrix} \right] \right]$$

- 输出张量形状 (1,6,4,5)，默认在“非 batch 维的最高维”上连接

$$\left[ \left[ \begin{bmatrix} 0. & 1. & 2. & 3. & 4. \\ 5. & 6. & 7. & 8. & 9. \\ 10. & 11. & 12. & 13. & 14. \\ 15. & 16. & 17. & 18. & 19. \end{bmatrix} \begin{bmatrix} 20. & 21. & 22. & 23. & 24. \\ 25. & 26. & 27. & 28. & 29. \\ 30. & 31. & 32. & 33. & 34. \\ 35. & 36. & 37. & 38. & 39. \end{bmatrix} \begin{bmatrix} 40. & 41. & 42. & 43. & 44. \\ 45. & 46. & 47. & 48. & 49. \\ 50. & 51. & 52. & 53. & 54. \\ 55. & 56. & 57. & 58. & 59. \end{bmatrix} \right] \right]$$

## axis

```
concatenationLayer = network.add_concatenation([inputT0, inputT0])
concatenationLayer.axis = 0 # 重设连接的维度，默认在倒数第三维（初始示例代码的 c 维）上连接
```

- 指定 axis=0（在最高维上连接），输出张量形状 (2,3,4,5)

$$\left[ \left[ \begin{bmatrix} 0. & 1. & 2. & 3. & 4. \\ 5. & 6. & 7. & 8. & 9. \\ 10. & 11. & 12. & 13. & 14. \\ 15. & 16. & 17. & 18. & 19. \end{bmatrix} \begin{bmatrix} 20. & 21. & 22. & 23. & 24. \\ 25. & 26. & 27. & 28. & 29. \\ 30. & 31. & 32. & 33. & 34. \\ 35. & 36. & 37. & 38. & 39. \end{bmatrix} \begin{bmatrix} 40. & 41. & 42. & 43. & 44. \\ 45. & 46. & 47. & 48. & 49. \\ 50. & 51. & 52. & 53. & 54. \\ 55. & 56. & 57. & 58. & 59. \end{bmatrix} \right] \right]$$

- 指定 axis=1（在次高维上连接），输出张量形状 (1,6,4,5)，结果与初始示例代码相同
- 指定 axis=2（在季高维上连接），输出张量形状 (1,3,8,5)

$$\left[ \begin{bmatrix} 0. & 1. & 2. & 3. & 4. \\ 5. & 6. & 7. & 8. & 9. \\ 10. & 11. & 12. & 13. & 14. \\ 15. & 16. & 17. & 18. & 19. \\ 0. & 1. & 2. & 3. & 4. \\ 5. & 6. & 7. & 8. & 9. \\ 10. & 11. & 12. & 13. & 14. \\ 15. & 16. & 17. & 18. & 19. \end{bmatrix} \begin{bmatrix} 20. & 21. & 22. & 23. & 24. \\ 25. & 26. & 27. & 28. & 29. \\ 30. & 31. & 32. & 33. & 34. \\ 35. & 36. & 37. & 38. & 39. \\ 20. & 21. & 22. & 23. & 24. \\ 25. & 26. & 27. & 28. & 29. \\ 30. & 31. & 32. & 33. & 34. \\ 35. & 36. & 37. & 38. & 39. \end{bmatrix} \begin{bmatrix} 40. & 41. & 42. & 43. & 44. \\ 45. & 46. & 47. & 48. & 49. \\ 50. & 51. & 52. & 53. & 54. \\ 55. & 56. & 57. & 58. & 59. \\ 40. & 41. & 42. & 43. & 44. \\ 45. & 46. & 47. & 48. & 49. \\ 50. & 51. & 52. & 53. & 54. \\ 55. & 56. & 57. & 58. & 59. \end{bmatrix} \right]$$

- 指定 axis=3（在叔高维上连接），输出张量形状 (1,3,4,10)

$$\left[ \begin{bmatrix} 0. & 1. & 2. & 3. & 4. & 0. & 1. & 2. & 3. & 4. \\ 5. & 6. & 7. & 8. & 9. & 5. & 6. & 7. & 8. & 9. \\ 10. & 11. & 12. & 13. & 14. & 10. & 11. & 12. & 13. & 14. \\ 15. & 16. & 17. & 18. & 19. & 15. & 16. & 17. & 18. & 19. \\ 20. & 21. & 22. & 23. & 24. & 20. & 21. & 22. & 23. & 24. \\ 25. & 26. & 27. & 28. & 29. & 25. & 26. & 27. & 28. & 29. \\ 30. & 31. & 32. & 33. & 34. & 30. & 31. & 32. & 33. & 34. \\ 35. & 36. & 37. & 38. & 39. & 35. & 36. & 37. & 38. & 39. \\ 40. & 41. & 42. & 43. & 44. & 40. & 41. & 42. & 43. & 44. \\ 45. & 46. & 47. & 48. & 49. & 45. & 46. & 47. & 48. & 49. \\ 50. & 51. & 52. & 53. & 54. & 50. & 51. & 52. & 53. & 54. \\ 55. & 56. & 57. & 58. & 59. & 55. & 56. & 57. & 58. & 59. \end{bmatrix} \right]$$