

# rewriting-caffe recode

- rewriting-caffe recode
  - 1. 编写CMakeLists.txt
  - 2. 文件功能说明
  - 3. caffe::string 的来源
  - 4. lint是C/C++ 强大的测试工具
  - 5. CPU\_ONLY 是怎么从cmake传给C++内部的
  - 6. device\_alternate.hpp
  - 7. common.hpp 解析
    - 7.1 boost::thread\_specific\_ptr是线程局部存储机制
  - 8. rng.hpp中的iterator\_trait
    - 8.1 value type
    - 8.2 difference\_type
    - 8.3 iterator\_category 迭代器的类别
  - 9.src/caffe/test/CMakeLists.txt中CACHE STRING
    - 9.1 定义
    - 9.2 两种变量的作用域原理和使用
      - 9.2.1 Normal Variables
      - 9.2.2 Cache Variables
  - 10. CMake 的function
  - 10. test\_caffe\_main.hpp 中的typedef typename
  - 11. 整个工程是怎么执行的呢？
  - 12. 使用cudaMallocHost的技巧
    - 12.1 cudaMallocHost, melloc, cudaMalloc
  - 13. SyncedMemory的设计很巧妙
  - 14. 在GPU模式下，make runtest caffe\_gpu\_memcpy无法被找到定义undefine reference
    - 14.1 其他undefine reference情况-C/C++混合编程导致的
    - 14.2 模板实例化和特例化的混了，导致undefine 错误
  - 15.Vscode 出现#include errors detected. Please update your includePath. IntelliSense
  - 16. test\_blob中的类型测试用例
  - 17. layerFactory
  - 18. NetStateRule
  - 18. 函数名后const
- 错误记录
  - 1. proto编译错误
  - 2. redefinition of kMaxBlobAxes

- [工程实现顺序](#)

按照一步一步的来实现caffe工程

## 1. 编写CMakeLists.txt

第一步编写CMakeLists.txt文档，有两个目的：

- 为了将整个工程结构有个大概了解，
- 在编写每个模块时，能跑通整个工程

## 2. 文件功能说明

```
#include "caffe/proto/caffe.pb.h"
```

包含caffe空间定义

如果不添加该文件的include，则会产生caffe没定义; 因为它是第一个生成的.h文件，且含有namespace caffe {}的caffe命名空间定义

## 3. caffe::string 的来源

```
是有namespace caffe{  
std::string  
}
```

产生的; 这个申明在common.hpp中，  
**因此开始先写common.hpp**

## 4. lint是C/C++ 强大的测试工具

lint检查C程序中潜在的错误，包括（但不限于）可疑的类型组合、未使用的变量、不可达的代码以及不可移植的代码。lint会产生一系列程序员有必要从头到尾仔细阅读的诊断信息。使用lint的好处是：1.它可以检查出被编译器漏掉的错误; 2.可以关联很多文件进行错误的检查和代码分析,具有较强大灵活性.lint可以检查的错误类型大体如下：

- 可能的空指针
- 在释放内存后使用了指向该内存的指针
- 赋值次序问题
- 拼写错误

- 被0除
- 失败的case语句(遗漏了break语句)
- 不可移植的代码(依赖了特定的机器实现)
- 宏参数没有使用圆括号
- 符号的丢失
- 异常的表达式
- 变量没有初始化
- 可疑的判断语句(例如,if(x=0))
- printf/scanf的格式检查

因此，代码中注释带有//NOLINT时，表示That's a comment. In this case, it's a comment designed to be read by a static analysis tool to tell it to shut up about this line. 让静态分析工具不要在意这一行。

## 5. CPU\_ONLY 是怎么从cmake传给C++内部的

在CaffeConfig.cmake.in中传入的, 具体怎么转化为#define CPU\_ONLY 的，需要考察cmake/Templates/CaffeConfig.cmake.in:53:set(Caffe\_CPU\_ONLY @CPU\_ONLY@)

ConfigGen.cmake 产生CaffeTargets.cmake, 这个文件中有：

```
set_target_properties(cafe PROPERTIES
  INTERFACE_COMPILE_DEFINITIONS "USE_LMDB;USE_LEVELDB;CPU_ONLY;USE_OPENCV"
  INTERFACE_INCLUDE_DIRECTORIES "/usr/include;/usr/local/include;/usr/local/include;/usr/include;/home/."
)
```

**CPU\_ONLY 在这传进去的**

**INTERFACE\_COMPILE\_DEFINITIONS：官方说明：**

List of public compile definitions requirements for a library.

Targets may populate this property to publish the compile definitions required to compile against the headers for the target. The target\_compile\_definitions() command populates this property with values given to the PUBLIC and INTERFACE keywords. Projects may also get and set the property directly.

When target dependencies are specified using target\_link\_libraries(), CMake will read this property from all target dependencies to determine the build properties of the consumer.

Contents of INTERFACE\_COMPILE\_DEFINITIONS may use “generator expressions” with the syntax \$<...>. See the [cmake-generator-expressions(7)] manual for available expressions. See the [cmake-buildsystem(7)] -manual for more on defining buildsystem properties.

**[cmake-generator-expressions(7)]---**: 也就是说这个INTERFACE\_COMPILE\_DEFINITIONS或者target\_compile\_definition的设置，会在编译的时候控制一些用宏定义为判断条件的代码是否编译编译，也就是控制条件性的定义

Generator expressions are evaluated during build system generation to produce information specific to each build configuration.

Generator expressions are allowed in the context of many target properties, such as LINK\_LIBRARIES, INCLUDE\_DIRECTORIES, COMPILE\_DEFINITIONS and others. They may also be used when using commands to populate those properties, such as target\_link\_libraries(), target\_include\_directories(), target\_compile\_definitions() and others.

**They enable conditional linking, conditional definitions used when compiling,** conditional include directories, and more. The conditions may be based on the build configuration, target properties, platform information or any other queryable information.

Generator expressions have the form \$<...>. To avoid confusion, this page deviates from most of the CMake documentation in that it omits angular brackets <...> around placeholders like condition, string, target, among others.

Generator expressions can be nested, as shown in most of the examples below.

扩展知识：

添加编译时的选项方法如下：

- add\_compile\_options() 可以给当前目录以及当前目录以下的目录的 sources 添加编译选项
- 2. target\_compile\_definitions：

```
target_compile_definitions(<target>
    <INTERFACE|PUBLIC|PRIVATE> [items1...]
    [<INTERFACE|PUBLIC|PRIVATE> [items2...] ...])
```

是给给定的 添加编译选项， 指的是由 add\_executable() 产生的可执行文件或 add\_library() 添加进来的库。 <INTERFACE|PUBLIC|PRIVATE> 指的是 [items...] 选项可以传播的范围， PUBLIC and INTERFACE 会传播 的 INTERFACE\_COMPILE\_DEFINITIONS 属性， PRIVATE and PUBLIC 会传播 的 COMPILE\_DEFINITIONS 属性。

- add\_definitions(-DFOO -DBAR ...) :可以给当前目录以及当前目录以下的目录的 sources 添加编译行命令。如 cmake . 这样的在命令行输入的命令。cmake的命令格式如下：

```
1 cmake [<options>] (<path-to-source> | <path-to-existing-build>)
2 cmake [(-D<var>=<value>)...] -P <cmake-script-file>
3 cmake --build <dir> [<options>] [-- <build-tool-options>...]
4 cmake -E <command> [<options>]
5 cmake --find-package <options>...
```

target\_compile\_options 与target\_compile\_definitions的区别

命名标签格式	说明
target_include_directories( [SYSTEM] [BEFORE] <INTERFACE PUBLIC PRIVATE> [items1...] [<INTERFACE PUBLIC PRIVATE> [items2...] ...])	Include的头文件的查找目录， 也就是Gcc的[-I dir...]选项
target_compile_definitions( <INTERFACE PUBLIC PRIVATE> [items1...][<INTERFACE PUBLIC PRIVATE> [items2...] ...])	通过命令行定义的宏变量， 也就是gcc的[-D macro[=defn]...]选项
target_compile_options( [BEFORE] <INTERFACE PUBLIC PRIVATE> [items1...] [<INTERFACE PUBLIC PRIVATE> [items2...] ...])	gcc其他的一些编译选项指定， 比如-fPIC

也就是说，target\_compile\_definitions用来定义编译时控制编译代码的宏定义的选项的，target\_compile\_options是编译规则的一些选项。 具体如下实例所示：

```

# gcc头文件查找目录，相当于-I选项，e.g -I/foo/bar
#CMAKE_SOURCE_DIR是cmake内置变量表示当前项目根目录
target_include_directories(test_elf
    PRIVATE
    ${CMAKE_SOURCE_DIR}
    ${CMAKE_SOURCE_DIR}/common
    ${CMAKE_SOURCE_DIR}/syscalls
)

# 编译的宏定义，e.g 相当于-D选项 e.g -Dmacro=defn
set(MONITOR_OMIT_BSS_INIT      "0")
set(MONITOR_OMIT_DATA_INIT     "0")
set(MONITOR_OMIT_T_CHECKS      "0")
target_compile_definitions(test_elf
    PRIVATE
    MONITOR_OMIT_BSS_INIT=${MONITOR_OMIT_BSS_INIT}
    MONITOR_OMIT_DATA_INIT=${MONITOR_OMIT_DATA_INIT}
    MONITOR_TRAP_NT_IRQS=${MONITOR_TRAP_NT_IRQS}
)

# 其他编译选项定义，e.g -fPIC
target_compile_options(test_elf
    PRIVATE
    -std=c99
    -Wall
    -Wextra
    -Werror
)

```

## 6. device\_alterate.hpp

配置 cuda的错误检查宏，以及配置block数目和threads数目

## 7. common.hpp 解析

这里有点绕，特别是Caffe类里面有个RNG，RNG这个类里面还有个Generator类  
在RNG里面会用到Caffe里面的Get()函数来获取一个新的Caffe类的实例（如果不存在的话）。  
然后RNG里面用到了Generator。Generator是实际产生随机数的。

#### (1) Generator类

该类有两个构造函数：

Generator()//用系统的熵池或者时间来初始化随机数

explicit Generator(unsigned int seed)// 用给定的种子初始化

#### (2) RNG类

RNG类内部有generator\_，generator\_是Generator类的实例

该类有三个构造函数：

RNG(); //利用系统的熵池或者时间来初始化RNG内部的generator\_

explicit RNG(unsigned int seed); // 利用给定的seed来初始化内部的generator\_

explicit RNG(const RNG&);// 用其他的RNG内部的generator\_设置为当前的generator\_

#### (3) Caffe类

1)含有一个Get函数，该函数利用Boost的局部线程存储功能实现

// Make sure each thread can have different values.

// boost::thread\_specific\_ptr是线程局部存储机制

// 一开始的值是NULL

static boost::thread\_specific\_ptr<Caffe> thread\_instance\_;

Caffe& Caffe::Get() {

if (!thread\_instance\_.get()) {// 如果当前线程没有caffe实例

thread\_instance\_.reset(new Caffe());// 则新建一个caffe的实例并返回

}

return \*(thread\_instance\_.get());

2)此外该类还有

SetDevice

DeviceQuery

mode

set\_mode

set\_random\_seed

solver\_count

set\_solver\_count

root\_solver

set\_root\_solver

等成员函数

3)内部还有一些比较技巧性的东西比如：

// CUDA: various checks for different function calls.

// 防止重定义cudaError\_t，这个以前在linux代码里面看过

// 实际上就是利用变量的局部声明

#define CUDA\_CHECK(condition) \

/\* Code block avoids redefinition of cudaError\_t error \*/ \

do { \

cudaError\_t error = condition; \

CHECK\_EQ(error, cudaSuccess) << " " << cudaGetErrorString(error); \

} while (0)

## 7.1 boost::thread\_specific\_ptr是线程局部存储机制

大多数函数都不是可重入的。这也就是说在某一个线程已经调用了函数时，如果你再调用同一个函数，那么这样是不安全的。一个不可重入的函数通过连续的调用来保存静态变量或者是返回一个指向静态数据的指针。举例来说，std::strtok就是不可重入的，因为它使用静态变量来保存要被分割成符号的字符串。

有两种方法可以让不可重用的函数变成可重用的函数。第一种方法就是改变接口，用指针或引用代替原先使用静态数据的地方。比方说，POSIX定义了strtok\_r，std::strtok中的一个可重入的变量，它用一个额外的char\*\*参数来代替静态数据。这种方法很简单，而且提供了可能的最佳效果。但是这样必须改变公共接口，也就意味着必须改代码。另一种方法不用改变公有接口，而是用本地存储线程（thread local storage）来代替静态数据（有时也被成为特殊线程存储，thread-specific storage）。

Boost线程库提供了智能指针boost::thread\_specific\_ptr来访问本地存储线程。每一个线程第一次使用这个智能指针的实例时，它的初值是NULL，所以必须要先检查这个它的只是否为空，并且为它赋值。Boost线程库保证本地存储线程中保存的数据会在线程结束后被清除。

List5是一个使用boost::thread\_specific\_ptr的简单例子。其中创建了两个线程来初始化本地存储线程，并有10次循环，每一次都会增加智能指针指向的值，并将其输出到std::cout上（由于std::cout是一个共享资源，所以通过互斥体进行同步）。main线程等待这两个线程结束后就退出。从这个例子输出可以明白的看出每个线程都处理属于自己的数据实例，尽管它们都是使用同一个boost::thread\_specific\_ptr。

### -----另一个解释

对于线程局部存储的概念，正如字面意思，每个变量在每个线程中都有一份独立的拷贝。通过使用线程局部存储技术，可以避免线程间的同步问题，并且不同的线程可以使用不同的日志设置。通过 Boost 库的智能指针 boost::thread\_specific\_ptr 来存取线程局部存储，每个线程在第一次试图获取这个智能指针的实例时需要对它进行初始化，并且线程局部存储的数据在线程退出时由 Boost 库来释放。

使用线程局部变量的多线程日志的优势：

使用 static 的线程局部变量很容易能实现线程级别的单例日志系统；

通过智能指针 boost::thread\_specific\_ptr 来存取线程局部存储的生成和清除工作简单方便；

使用线程局部变量很容易实现对于一个已有的单例日志系统进行多线程支持的改造，并且不用改动任何原来的日志接口；

## 8. rng.hpp中的iterator\_trait

这个类的作用是得到某个迭代器的相关信息。比如 iterator\_traits<vector::iterator>::difference\_type。



这个API用于返回模板的具体相关信息，包括迭代器每一项之间的距离difference type， 迭代器每一项的值。iterator\_traits负责萃取迭代器的特性。

- 1、value type 用来表示迭代器所指对象的型别；
- 2、difference type 用来表示两个迭代器之间的距离；
- 3、reference 为引用类型；
- 4、pointer 为指针类型；
- 5、iterator\_category 表明迭代器的类型；

## 8.1 value type

用来表示迭代器所指对象的型别；

## 8.2 difference\_type

用来表示两个迭代器之间的距离；头尾之间的距离为容器的最大容量。

difference\_type是一种常用的迭代器型别，用来表示两个迭代器之间的距离，因此它也可以用来表示一个容器的最大容量，因为对于连续空间的容器而言，头尾之间的距离就是其最大容量。

## 8.3 iterator\_category 迭代器的类别

返回迭代器类别，

迭代器被分为五类：

- 1、Input Iterator：这种迭代器所指对象，不允许外界改变，只读（read only）；
- 2、Output Iterator：唯写（write only）；
- 3、Forward Iterator：允许「写入型」算法（例如 replace()）在此种迭代器所形成的区间上做读写动作；
- 4、Bidirectional Iterator：可双向移动。某些算法需要逆向走访某个迭代器区间（例如逆向拷贝某范围内的元素），就可以使用 Bidirectional Iterators；
- 5、Random Access Iterator：前四种迭代器都只供应一部份指标算术能力（前3种支持 operator++，第4种再加上 operator--），第5种则涵盖所有指标算术能力，包括 p+n, p-n, p[n], p1-p2, p1<p2.

```
input_iterator_tag
output_iterator_tag
forward_iterator_tag
bidirectional_iterator_tag
random_access_iterator_tag
```

## 9.src/caffe/test/CMakeLists.txt中CACHE STRING

CMake 变量包含 Normal Variables、Cache Variables。

## 9.1 定义

- **Normal Variables**：通过 `set( ... [PARENT_SCOPE])` 这个命令来设置的变量就是 Normal Variables。例如 `set(MY_VAL "666")`，此时 MY\_VAL 变量的值就是 666。
- **Cache Variables**：通过 `set( ... CACHE [FORCE])` 这个命令来设置的变量就是 Cache Variables。例如 `set(MY_CACHE_VAL "666" CACHE STRING INTERNAL)`，此时 MY\_CACHE\_VAL 就是一个 CACHE 变量。

## 9.2 两种变量的作用域原理和使用

### 9.2.1 Normal Variables

作用域属于整个 CMakeLists.txt 文件，当该文件包含了 `add_subdirectory()`、`include()`、`macro()`、`function()` 语句时，会出现两种不同的效果：

- 包含 `add_subdirectory()`、`function()`。（本质是值拷贝）：假设，我们在工程根目录 CMakeLists.txt 文件中使用 `add_subdirectory(src)` 包含另一个 src 目录，在 src 目录中有另一个 CMakeLists.txt 文件。以根目录 CMake 文件为父目录，src 目录为子目录，此时子目录 CMake 文件会拷贝一份父目录文件的 Normal 变量。需要说明的是，我们在子目录中如果想要修改父目录 CMake 文件包含的 Normal 变量。必须通过 `set(... PARENT_SCOPE)` 的方式，再 src 目录下需要写 `set(MY_VAL "777" PARENT_SCOPE)` # 修改处。
- 包含 `include()`、`macro()`（本质有点类似 c 中的 `#include` 预处理含义），会直接更改

### 9.2.2 Cache Variables

相当于一个全局变量，我们在同一个 cmake 工程中都可以使用。Cache 变量有以下几点说明

- Cache 变量 `CMAKE_INSTALL_PREFIX` 默认值是 `/usr/local` (可以在生成的 CMakeCache.txt 文件中查看)，这时候如果我们 在某个 CMakeLists.txt 中，仍然使用 `set(CMAKE_INSTALL_PREFIX "/usr")`，那么此时我们 install 的时候，CMake 以后面的 `/usr` 作为 `CMAKE_INSTALL_PREFIX` 的值，这是因为 CMake 规定，有一个与 Cache 变量同名的 Normal 变量出现时，后面使用这个变量的值都是以 Normal 为准，如果没有同名的 Normal 变量，CMake 才会自动使用 Cache 变量。
- 所有的 Cache 变量都会出现在 CMakeCache.txt 文件中。这个文件是我们键入 `cmake .` 命令后自动出现的文件。打开这个文件发现，CMake 本身会有一些默认的全局 Cache 变量。例如：`CMAKE_INSTALL_PREFIX`、`CMAKE_BUILD_TYPE`、`CMAKE_CXX_FLAGS` 等等。可以自行查看。当然，我们自己定义的 Cache 变量也会出现在这个文件中。Cache 变量定义格式为 `set(CACHE STRING INTERNAL)`。这里的 STRING 可以替换为 `BOOL FILEPATH PATH`，但是要根据前面 value 类型来确定。

- 修改 Cache 变量。可以通过 set( CACHE INTERNAL FORCE)，另一种方式是直接在终端中使用 cmake -D var=value ..来设定默认存在的CMake Cache 变量。

## 10. CMake 的function

```
# Filter out all files that are not included in selected list
# Usage:
#caffe_leave_only_selected_tests(<filelist_variable> <selected_list>)

function(caffe_leave_only_selected_tests file_list)
    # ARGV 表示<selected_list>的传入， ARGV是后面参数的一个占位符，
    # 如果用caffe_leave_only_selected_tests(var1, var2)，那么ARGV就是var2
    if(NOT ARGV)
        return() # blank list means leave all
    endif()
    # message(STATUS "caffe_leave_only_selected_tests - ARGV: ${ARGV}")
    string(REPLACE " " ";" _selected ${ARGV})
    list(APPEND _selected caffe_main)

    # message(STATUS "caffe_leave_only_selected_tests, file_list: ${file_list}")
    set(result "")
    foreach(f ${file_list})
        get_filename_component(name ${f} NAME_WE)
        string(REGEX REPLACE "^test_" "" name ${name})
        list(FIND _selected ${name} _index)
        if(NOT _index EQUAL -1)
            list(APPEND result ${f})
        endif()
    endforeach()
    set(${file_list} ${result} PARENT_SCOPE)
endfunction()
```

## 10. test\_caffe\_main.hpp 中的typedef typename

```
typedef typename TypeParam::Dtype Dtype;
```

如果给模板类型添加别名时，需要加上typename

如果不加这个关键字，编译器就不知道TypeParam::Dtype到底是个什么东西？可能是静态成员变量，也有可能是静态成员函数，也有可能是内部类。

加上这个关键字等于手动告诉编译器：TypeParam::Dtype就是一个类型。

也就是说 模板类型在实例化之前，编译器并不知道

typedef创建了存在类型的别名，而typename告诉编译器TypeParam::Dtype是一个类型而不是一个成

员。

## 11. 整个工程是怎么执行的呢？

- 先给caffe 目标，所需要的所有源代码，srcs，然后这样在生成该target时，会将所有的srcs中的.cpp, .cu, 进行编译，这些文件中include的头文件会被找到，然后作为这些源文件的依赖进行编译，这个会自动生成makefile语法中cxxxx.o:xx.hpp,xxx.h的格式
- target\_link\_libraries给出所有需要的库
- 给出所有需要的 .hpp, .h 文件
- target\_compile\_definitions 给出一些定义这个是用来控制程序中像'#ifdef CPU\_ONLY'这样的宏判断语句的；如下面配置，\${Caffe\_DEFINITIONS}包含-DCPU\_ONLY，-DUSE\_MKL等， 这样在编译程序的时候，或根据程序中#ifndef CPU\_ONLY 来控制编译哪些代码？

```
# Caffe_DEFINITIONS变量设置
set(Caffe_DEFINITIONS "")
cmake/Dependencies.cmake:63: list(APPEND Caffe_DEFINITIONS PUBLIC -DUSE_LMDB)
cmake/Dependencies.cmake:65: list(APPEND Caffe_DEFINITIONS PRIVATE -DALLOW_LMDB_NOLOC
cmake/Dependencies.cmake:74: list(APPEND Caffe_DEFINITIONS PUBLIC -DUSE_LEVELDB)
cmake/Dependencies.cmake:93: list(APPEND Caffe_DEFINITIONS PUBLIC -DCPU_ONLY)
cmake/Dependencies.cmake:112: list(APPEND Caffe_DEFINITIONS PUBLIC -DUSE_OPENCV)
cmake/Dependencies.cmake:132: list(APPEND Caffe_DEFINITIONS PUBLIC -DUSE_MKL)
cmake/Dependencies.cmake:141: list(APPEND Caffe_DEFINITIONS PUBLIC -DUSE_ACCELERATE
cmake/Dependencies.cmake:185: list(APPEND Caffe_DEFINITIONS PRIVATE -DWITH_PYTHON_L
cmake/ConfigGen.cmake:36: list(APPEND Caffe_DEFINITIONS -DUSE_HDF5)
cmake/Cuda.cmake:257: list(APPEND Caffe_DEFINITIONS PUBLIC -DUSE_CUDNN)
```

- set\_target\_properties() 设置一些属性

```

# 主要的工程编译程序
# 因为一些库和anaconda3/libs冲突了，它选的是anaconda3里面的库
add_library(caffe ${srcs})
caffe_default_properties(caffe)
message(STATUS "add lib caffe Caffe_LINKER_LIBS: ${Caffe_LINKER_LIBS} ")
target_link_libraries(caffe ${Caffe_LINKER_LIBS})
target_include_directories(caffe ${Caffe_INCLUDE_DIRS}
                            PUBLIC
                            $<BUILD_INTERFACE:${Caffe_INCLUDE_DIR}>
                            $<INSTALL_INTERFACE:include>)
message(STATUS "caffe definition: ----- ${Caffe_DEFINITIONS}, ${Caffe_COMPILE_OPTIONS} ")
target_compile_definitions(caffe ${Caffe_DEFINITIONS})
if(Caffe_COMPILE_OPTIONS)
    message(STATUS "add lib caffe Caffe_COMPILE_OPTIONS:  ${Caffe_COMPILE_OPTIONS}")
    target_compile_options(caffe ${Caffe_COMPILE_OPTIONS})
endif()
set_target_properties(caffe PROPERTIES
    VERSION    ${CAFFE_TARGET_VERSION}
    SOVERSION  ${CAFFE_TARGET_SOVERSION}
)

```

## 12. 使用cudaMallocHost的技巧

syncedMemory 中有个技巧：CaffeMallocHost 中判断ptr指针是否为GPU指针，如果是则用 cudaMallocHost生成pinned memory这样高效；如果不是则malloc进行分配。这样能高效执行GPU->CPU的传输。而且pinned mem的设置主要对Memcpy函数的高效传输有用。

### 12.1 cudaMallocHost, melloc, cudaMalloc

分页（英语：Paging），是一种操作系统里存储器管理的一种技术，可以使电脑的主存可以使用存储在辅助存储器中的数据。操作系统会将辅助存储器（通常是磁盘）中的数据分区成固定大小的区块，称为“页”（pages）。当不需要时，将分页由主存（通常是内存）移到辅助存储器；当需要时，再将数据取回，加载主存中。相对于分段，分页允许存储器存储于不连续的区块以维持文件系统的整齐。分页是磁盘和内存间传输数据块的最小单位。

我们用cudaMalloc()为GPU分配内存,用malloc()为CPU分配内存.除此之外,CUDA还提供了自己独有的机制来分配host内存:cudaHostAlloc(). 这个函数和malloc的区别是什么呢？

malloc()分配的标准的,可分页的主机内存(上面有解释到),而cudaHostAlloc()分配的是页锁定的主机内存,也称作固定内存pinned memory,或者不可分页内存,它的一个重要特点是操作系统将不会对这块内存分页并交换到磁盘上,从而保证了内存始终驻留在物理内存中.也正因为如此,操作系统能够安全地使某个应用程序访问该内存的物理地址,因为这块内存将不会被破坏或者重新定位.

由于GPU知道内存的物理地址,因此就可以使用DMA技术来在GPU和CPU之间复制数据.当使用可分页的内存进行复制时(使用malloc),CUDA驱动程序仍会通过dram把数据传给GPU,这时复制操作会执行两遍,第一遍从可分页内存复制一块到临时的页锁定内存,第二遍是再从这个页锁定内存复制到GPU上.当从可分页内存中执行复制时,复制速度将受限制于PCIE总线的传输速度和系统前端速度相对较低的一方.在某些系统中,这些总线在带宽上有着巨大的差异,因此当在GPU和主机之间复制数据时,这种差异会使页锁定主机内存比标准可分页的性能要高大约2倍.即使PCIE的速度于前端总线的速度相等,由于可分页内存需要更多一次的CPU参与复制操作,也会带来额外的开销.

当我们在调用cudaMemcpy(dest, src, ...)时,程序会自动检测dest或者src是否为Pinned Memory,若不是,则会自动将其内容拷入一不可见的Pinned Memory中,然后再进行传输.可以手动指定Pinned Memory,对应的API为:cudaHostAlloc(address, size, option)分配地址, cudaFreeHost(pointer)释放地址.注意,所谓的Pinned Memory都是在Host端的,而不是Device端。

有的人看到这里,在写代码的过程中把所有的malloc都替换成cudaHostAlloc()这样也是不对的.

固定内存是一把双刃剑.当时使用固定内存时,虚拟内存的功能就会失去,尤其是,在应用程序中使用每个页锁定内存时都需要分配物理内存,而且这些内存不能交换到磁盘上.这将会导致系统内存会很快被耗尽,因此应用程序在物理内存较少的机器上会运行失败,不仅如此,还会影响系统上其他应用程序的性能.综上所述,建议针对cudaMemcpy()调用中的源内存或者目标内存,才使用页锁定内存,并且在不在使用他们的时候立即释放,而不是在应用程序关闭的时候才释放.我们使用下面的测试实例

## 13. SyncedMemory的设计很巧妙

有四个状态, 用to\_cpu, to\_gpu进行状态之间的转移判断; 利用cpu\_data, mutable\_cpu\_data, gpu\_data, mutable\_gpu\_data进行不同需求的数据读取; 也就是一个任务要分离出需要的接口函数, 以及确定好那些是内部的固定的功能函数如to\_cpu, to\_gpu;

## 14. 在GPU模式下, make runtest caffe\_gpu\_memcpy无法被找到定义undefine reference

可能链接库的顺序不对, 而且添加了caffe::GlobalInit(&argc,&argv) make就出该错误;  
真正原因: 是.hpp中的声明, 和.cu中的定义不一样导致的

```
# declare in math_functions.hpp
void caffe_gpu_memcpy(const size_t n, const void* x, const void* y); // wrong
// 应该将 第二个const去掉
void caffe_gpu_memcpy(const size_t n, const void* x, void* y);

# define in math_functions.cu
void caffe_gpu_memcpy(const size_t n, const void* x, void* y){
    if (x != y){
        // x ---> y
        CUDA_CHECK(cudaMemcpy(y, x, n, cudaMemcpyDefault)); // NOLINT(caffe/alt_fn)
    }
}
```

所以以后，如果出现这种提示，先找书写格式上的错误

## 14.1 其他undefine reference情况-C/C++混合编程导致的

因为在cblas.h中的实现都是c编译器编译的，所以要在cblas.h 外加上extern "C" {};

```
extern "C" {
    #include <cblas.h>
}
```

## 14.2 模板实例化和特例化的混了，导致undefine 错误

实例化：

```
template <typename Dtype>
Dtype caffe_cpu_dot(const int n, const Dtype* x, const Dtype* y) {
    return caffe_cpu_strided_dot(n, x, 1, y, 1);
}

template
float caffe_cpu_dot<float>(const int n, const float* x, const float* y);

template
double caffe_cpu_dot<double>(const int n, const double* x, const double* y);
```

而错误是把实例化的方式写成特例化如添加了<> 符号

特例化

```

template <typename Dtype>
Dtype caffe_cpu_dot(const int n, const Dtype* x, const Dtype* y) {
    return caffe_cpu_strided_dot(n, x, 1, y, 1);
}
template <>
float caffe_cpu_dot<float>(const int n, const float* x, const float* y);

template <>
double caffe_cpu_dot<double>(const int n, const double* x, const double* y);

```

这是错误的，因为特例化需要定义的也就是需要具体实现的。而这直接加了一个分号“;”，所以没有它的实现，因此报出

undefine reference错误。

## 15.Vscode 出现#include errors detected. Please update your includePath. IntelliSense

shift+ctrl+p, 然后输入C/C++ 然后找C/C++ configure file (JSON); 然后配置如下

```

{
    "configurations": [
        {
            "name": "Linux",
            "includePath": [
                # 一些.h 如driver_types.h 无法被加入，但是能跳转过去； 需要添加该文件所在路径，因此添加了一下可能会需要
                "/usr/include/**",
                "/usr/local/cuda/include/**",
                "/usr/local/include/**",
                "${workspaceFolder}/**"
            ],
            "defines": [],
            "compilerPath": "/usr/bin/gcc",
            "cStandard": "c11",
            "cppStandard": "c++11",
            "intelliSenseMode": "gcc-x64"
        }
    ],
    "version": 4
}

```

## 16. test\_blob中的类型测试用例

首先有个testing::Test子类，用于检测多设备上的代码测试



```

template <typename TypeParam>
class MultiDeviceTest : public ::testing::Test {
public:
    typedef typename TypeParam::Dtype Dtype;
protected:
    MultiDeviceTest() {
        Caffee::set_mode(TypeParam::device);
    }
    virtual ~MultiDeviceTest() {}
};
// 注意TypeParam::device 模板类型有device属性，

// 根据下面的实现代码可以发现： TypeParam模板类型是一个结构体；
// 模板类型可以是一个模板类型的结构体，所以，实例化时，仍然还有Dtype的模板类型，用于表示模板结构体的类型
//////// CPU device的实例化-----
template <typename TypeParam>
struct CPUDevice {
    typedef TypeParam Dtype;
    static const Caffee::Brew device = Caffee::CPU;
};
// 实例化MultiDeviceTest 用结构体CPUDevice<Dtype>
template <typename Dtype>
class CPUDeviceTest:public MultiDeviceTest<CPUDevice<Dtype> > {
};
// 类型测试
typedef ::testing::Types<float, double> TestDtypes;

// 设备结构体类型的测试
typedef ::testing::Types<CPUDevice<float>, CPUDevice<double>,
GPUDevice<float>, GPUDevice<double> >
TestDtypesAndDevices;

```

在test\_blob.cpp中

```

// 类型测试 的使用,
typedef ::testing::Types<float, double> TestDtypes;
// 第一步： 定义测试类
template <typename Dtype>
class BlobSimpleTest : public ::testing::Test {
protected:
BlobSimpleTest()
    : blob_(new Blob<Dtype>()),
      blob_preshaped_(new Blob<Dtype>(2, 3, 4, 5)) {}
virtual ~BlobSimpleTest() { delete blob_; delete blob_preshaped_; }
Blob<Dtype>* const blob_;
Blob<Dtype>* const blob_preshaped_;
};

// 第二步： 类型测试变量，配置给测试类
// 对测试类进一步的设置， 设置他要测试的类型
TYPED_TEST_CASE(BlobSimpleTest, TestDtypes); // BlobSimpleTest

// 用TYPED_TEST进行测试
TYPED_TEST(BlobSimpleTest, TestInitialization){
    EXPECT_TRUE(this->blob_); // 用这个用例类的内部成员变量
    EXPECT_TRUE(this->blob_preshaped_);
    EXPECT_EQ(this->blob_preshaped_->num(), 2);
    EXPECT_EQ(this->blob_preshaped_->channels(), 3);
    EXPECT_EQ(this->blob_preshaped_->height(), 4);
    EXPECT_EQ(this->blob_preshaped_->width(), 5);
    EXPECT_EQ(this->blob_preshaped_->count(), 120);
    EXPECT_EQ(this->blob_->num_axes(), 0);
    EXPECT_EQ(this->blob_->count(), 0);
}
////////***** 设备测试也同等类似书写方式

```

设备类型的测试（类型是结构体）

```

template <typename TypeParam>
class BlobMathTest: public MultiDeviceTest<TypeParam> {
    typedef typename TypeParam::Dtype Dtype; // typename 将模板表示为类型
protected:
    BlobMathTest():
        blob_(new Blob<Dtype>(2,3,4,5)),
        epsilon_(1e-6){}
    virtual ~BlobMathTest() {
        delete blob_;
    }
    Blob<Dtype>* const blob_; // 指针内容不可变
    Dtype epsilon_;
};

/**
 * typedef ::testing::Types<CPUDevice<float>,
 * CPUDevice<double>,
 * GPUDevice<float>, GPUDevice<double> >
 * TestDtypesAndDevices;
 */
TYPED_TEST_CASE(BlobMathTest, TestDtypesAndDevices);

TYPED_TEST(BlobMathTest, TestsumOfSquares) {
    // TypeParam 是MultiDeviceTest的模板变量， 被实例化成GPUDevice和CPUDevice 结构体
    // 也就是TypeParam实际上表示为结构体变量
    /**
    template <typename TypeParam> // 此处的TypeParam表示float或double
    struct GPUDevice {
        typedef TypeParam Dtype;
        static const Caffe::Brew device = Caffe::GPU;
    };或

    template <typename TypeParam>
    struct CPUDevice {
        typedef TypeParam Dtype;
        static const Caffe::Brew device = Caffe::CPU;
    };
    */
    // 注意每个测试用例都可以TypeParam的方式访问测试用例的类（如BlobMathTest-继承自testing::Test）
    typedef typename TypeParam::Dtype Dtype;

    // uninitialized blob should have sum of squares == 0
    EXPECT_EQ(0, this->blob_->sumsq_data());
    EXPECT_EQ(0, this->blob_->sumsq_diff());

    FillerParameter filler_param;
    filler_param.set_min(-3);
    filler_param.set_max(3);
    UniformFiller<Dtype> filler(filler_param);
    filler.Fill(this->blob_);

```

```

Dtype expected_sumsq = 0;
const Dtype * data = this->blob_->cpu_data();
for (int i = 0; i < this->blob_->count(); ++i){
    expected_sumsq += data[i] * data[i];
}

// Do a mutable access on the current device,
// so that the sumsq computation is done on that device.
// (Otherwise, this would only check the CPU sumsq implementation.)
switch (TypeParam::device)
{
case Caffe::CPU:
    // mutable的函数调用包含数据在不同设备上的传输过程，这样能测试更全面
    this->blob_->mutable_cpu_data();
    break;
case Caffe::GPU:
    this->blob_->mutable_gpu_data();
    break;
default:
    LOG(FATAL) << "Unknown device: " << TypeParam::device;
}
EXPECT_NEAR(expected_sumsq, this->blob_->sumsq_data(),
    this->epsilon_ * expected_sumsq);
EXPECT_EQ(0, this->blob_->sumsq_diff());

// check sumsq_diff too

}

```

单纯的类型测试的使用：test\_blob.cpp

```

// test_caffe_main 定义要测试的类型
typedef ::testing::Types<float, double> TestDtypes;

// 先实现一个测试子类
template <typename Dtype>
class BlobSimpleTest : public ::testing::Test {
protected:
BlobSimpleTest()
    : blob_(new Blob<Dtype>()),
      blob_preshaped_(new Blob<Dtype>(2, 3, 4, 5)) {}
virtual ~BlobSimpleTest() { delete blob_; delete blob_preshaped_; }
Blob<Dtype>* const blob_;
Blob<Dtype>* const blob_preshaped_;
};
// 配置将该类的类型测试case： 使用了上面定义的
TYPED_TEST_CASE(BlobSimpleTest, TestDtypes); // BlobSimpleTest

///// TYPED_TEST进行测试
TYPED_TEST(BlobSimpleTest, TestReshape){
    this->blob_->Reshape(2, 3, 4, 5);
    EXPECT_EQ(this->blob_->num(), 2);
    EXPECT_EQ(this->blob_->channels(), 3);
    EXPECT_EQ(this->blob_->height(), 4);
    EXPECT_EQ(this->blob_->width(), 5);
    EXPECT_EQ(this->blob_->count(), 120);
}

```

## 17. layerFactory

该类是一个注册工厂模式， 用map实现字符串到创建实例函数指针的映射

## 18. NetStateRule

其中的stage还不知道干什么用的？

```
message NetStateRule {
    // Set phase to require the NetState have a particular phase (TRAIN or TEST)
    // to meet this rule.
    // 设置这一层在那个阶段被include或者被exclude
    optional Phase phase = 1;

    // Set the minimum and/or maximum levels in which the layer should be used.
    // Leave undefined to meet the rule regardless of level.
    // 设置等级
    optional int32 min_level = 2;
    optional int32 max_level = 3;

    // Customizable sets of stages to include or exclude.
    // The net must have ALL of the specified stages and NONE of the specified
    // "not_stage"s to meet the rule.
    // (Use multiple NetStateRules to specify conjunctions of stages.)
    // 这个参数还不知道干什么用
    repeated string stage = 4;
    repeated string not_stage = 5;
}
```

具体作用说明：

作用：StateMeetsRule()中net的state是否满足NetStateRule

用构造net时的输入phase/level/stage与prototxt中各层的规则(include/exclude)比较, 决定本层是否要包含

1. Phase: train, test, 比如train的layer不适用于test
2. Min\_level: 本层level不小于min\_level, 则满足包含条件
3. Max\_level: 本层level不大于max\_level, 则满足包含条件
4. Stage: stage能在NetStateRule::stage中找到, 则包含本层
5. Non\_stage: stages能在NetStateRule::non\_stage中找到, 则排除本层

解释

在caffe中, 所有参数结构定义在caffe.proto中, 由protobuf的protoc.exe生成caffe.pb.c及caffe.pb.h, 在定义net网络结构的prototxt文件中往往会定义某层的include/exclude参数, 以确定该层网络是否要包含在某顾名思义, include表示如果在构造net时如果满足include的条件, 本层就包含在net中; exclude表示在构造net

管理这个被读取后的include还是exclude参数的, 就是caffe.proto中的NetStateRule类, 类中有phase、min

```
Net<Dtype>::Net(const string& param_file, Phase phase, const int level, const vector<str
```

对于包含include参数的层: 如果满足min\_level<level<max\_level 或 stages中任意一个元素能在NetStateRule::stage中找到, 则包含本层

对于包含exclude参数的层: 如果满足min\_level<level<max\_level 或 stages中任意一个元素能在NetStateRule::non\_stage中找到, 则排除本层

当然如果是在NetStateRule::non\_stage中找到, 结果正好相反, 看下面的例子,

```
layer {
  name: "mnist"
  type: "Data"
  top: "data"
  top: "label"
  include {
    phase: TEST
    not_stage: "predict" # 在 predict 时过滤掉这一层
  }
  transform_param {
    scale: 0.00390625
  }
  data_param {
    source: "examples/mnist/mnist_test_lmdb"
    batch_size: 100
    backend: LMDB
  }
}

# 增加 deploy 的输入层
layer {
  name: "data"
  type: "Input"
  top: "data"
  input_param { shape: { dim: 1 dim: 1 dim: 28 dim: 28 } }
  exclude {
    phase: TEST
    stage: "predict" # 在 predict 时不加上这一层
  }
}
```

```
}
```

如果想进一步了解对参数进行过滤有什么实际用处，我推荐这篇文章< Caffe 神经网络配置 - All in one netw

[https://yangwenbo.com/articles/caffe-net-config-all-in-one.html?utm\\_source=tuicool&utm\\_r](https://yangwenbo.com/articles/caffe-net-config-all-in-one.html?utm_source=tuicool&utm_r)

stage的使用：

Python:

```
net = caffe.Net("train_val_deploy.prototxt", caffe.TEST, stages=['predict'],
                weights="iter_N.caffemodel")
```

C++:

```
caffe::vector<caffe::string> stages;
stages.push_back("predict");
caffe::Net *net = new caffe::Net("train_val_deploy.prototxt", caffe::TEST, 0, &stages);
```

## 18. 函数名后const

```
/**
 * @brief Returns true if the layer requires an equal number of bottom and
 *        top blobs.
 *
 * This method should be overridden to return true if your layer expects an
 * equal number of bottom and top blobs.
 * 函数名后面的const 表示给函数不能改变类的成员，这样会安全很多，只要这个函数不是用来改成员
 */
virtual inline bool EqualNumBottomTopBlobs() const { return false; }
```

## 错误记录

### 1. proto编译错误

```
[ 14%] Building CXX object src/caffe/CMakeFiles/caffeproto.dir/__/__/include/caffe/protoc
c++: fatal error: no input files
compilation terminated.
/bin/sh: 1: -fPIC: not found
/bin/sh: 1: -Wall: not found
src/caffe/CMakeFiles/caffeproto.dir/build.make:74: recipe for target 'src/caffe/CMakeFil
make[2]: *** [src/caffe/CMakeFiles/caffeproto.dir/__/__/include/caffe/proto/caffe.pb.cc.
CMakeFiles/Makefile2:183: recipe for target 'src/caffe/CMakeFiles/caffeproto.dir/all' fa
make[1]: *** [src/caffe/CMakeFiles/caffeproto.dir/all] Error 2
Makefile:127: recipe for target 'all' failed
make: *** [all] Error 2
```

查看 `protoc --version` 2.6.1, 不是版本的问题



根据提示可以发现：

-fPIC not found

-Wall not found

查看错误处代码：

```
src/caffe/CMakeFiles/caffeproto.dir/__/__/include/caffe/proto/caffe.pb.cc.o: include/caf
    @$(CMAKE_COMMAND) -E cmake_echo_color --switch=$(COLOR) --green --progress-dir=/
    cd /home/lshn/Projects/rewrite_caffe/build/src/caffe && /usr/bin/c++ $(CXX_DEF
```

1) 去掉 -fPIC, -Wall 编译成功

2) c++ 换成g++ 也不行

通过打印发现，\${CXX\_FLAGS}的内容是 ";-fPIC;-Wall"，而一般这两个编译选项是不带";"分号的，因此赋值时出错；而CXX\_FLAGS是make的系统变量，它是由CMAKE\_CXX\_FLAGS的值生成的。而根目录下的CMake

```
set(CMAKE_CXX_FLAGS "${CMAKE_CXX_FLAGS}" -fPIC -Wall)
```

对照原来的caffe的CMakeLists.txt是

```
set(CMAKE_CXX_FLAGS "${CMAKE_CXX_FLAGS} -fPIC -Wall")
```

所以修改后错误消失

## 2. redefinition of kMaxBlobAxes

- 这是在编译的时候出错，（因为错误说明指向的是.cpp/.h文件所以是编译出错，而不是链接出错）
- 编译出错需要再blob中添加ifndef，而ifndef添加的不太对导致了错误。

## 工程实现顺序

1. caffe 一半，然后需要common.hpp
2. common.hpp 有需要cudnn, device\_alternate,
3. 写common的测试代码test\_common.cpp，而它又需要syncdmem.hpp和math\_functions.hpp
4. blob.hpp, blob.cpp, filler.hpp（对weight blob的填充） test\_blob.cpp