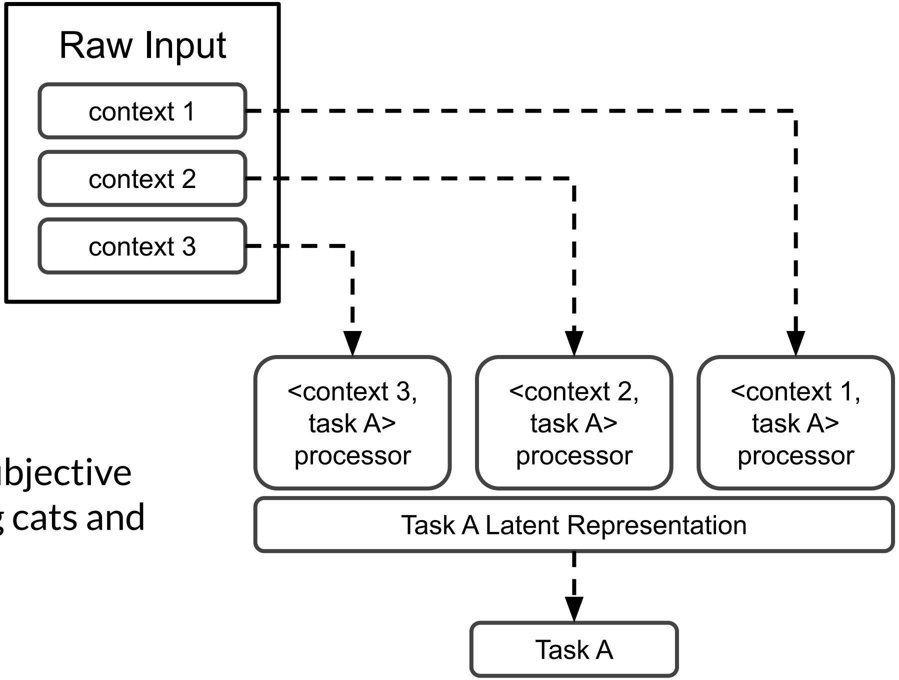
Meaning-Preserving Continual Learning v1

Meaning

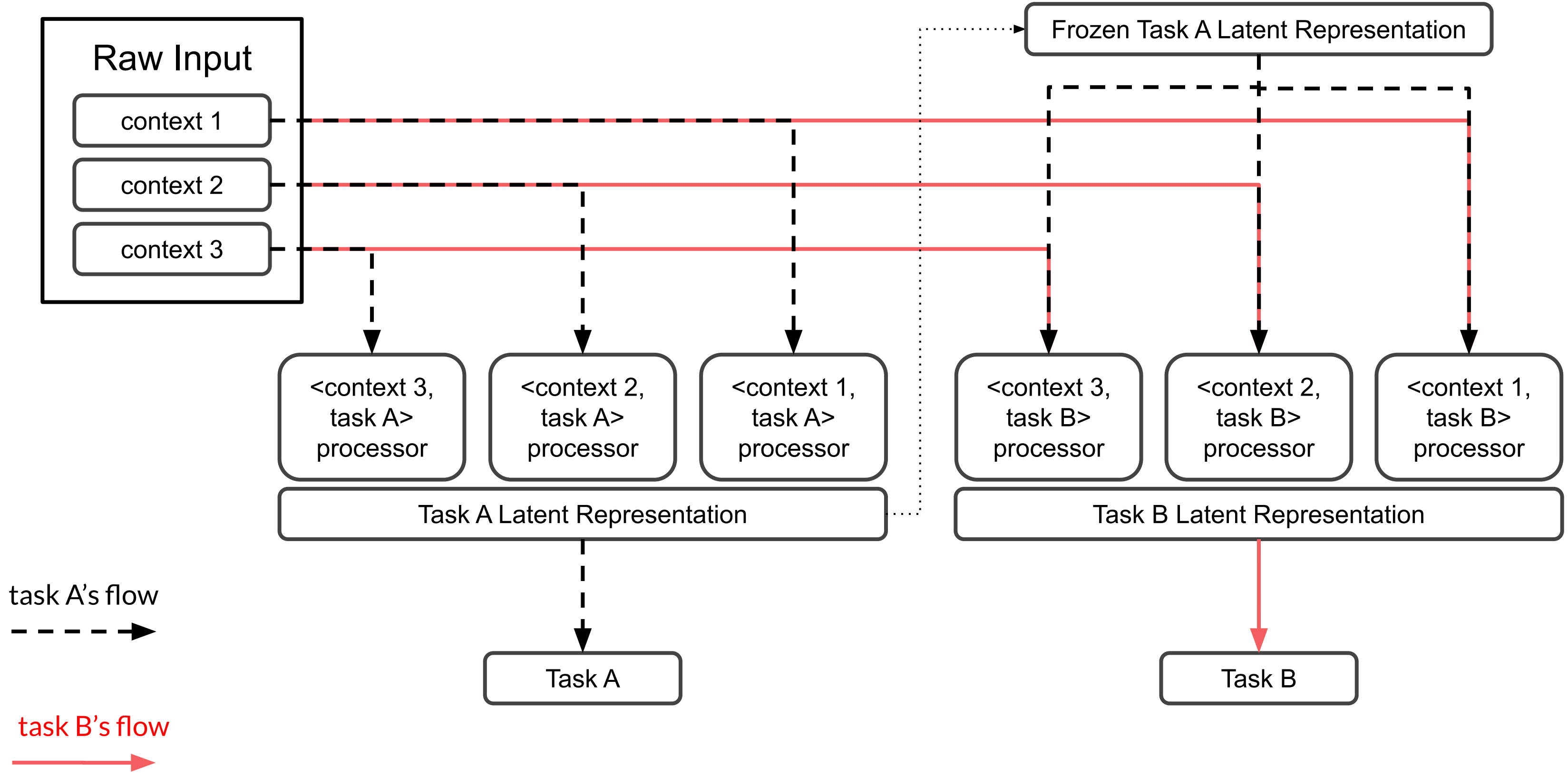

Raw Input
context 1
context 2
context 3
<context 3, task A> processor
<context 2, task A> processor
<context 1, task A> processor
Task A Latent Representation
Task A

Task A is expressed in subjective human terms, e.g. telling cats and dogs apart.

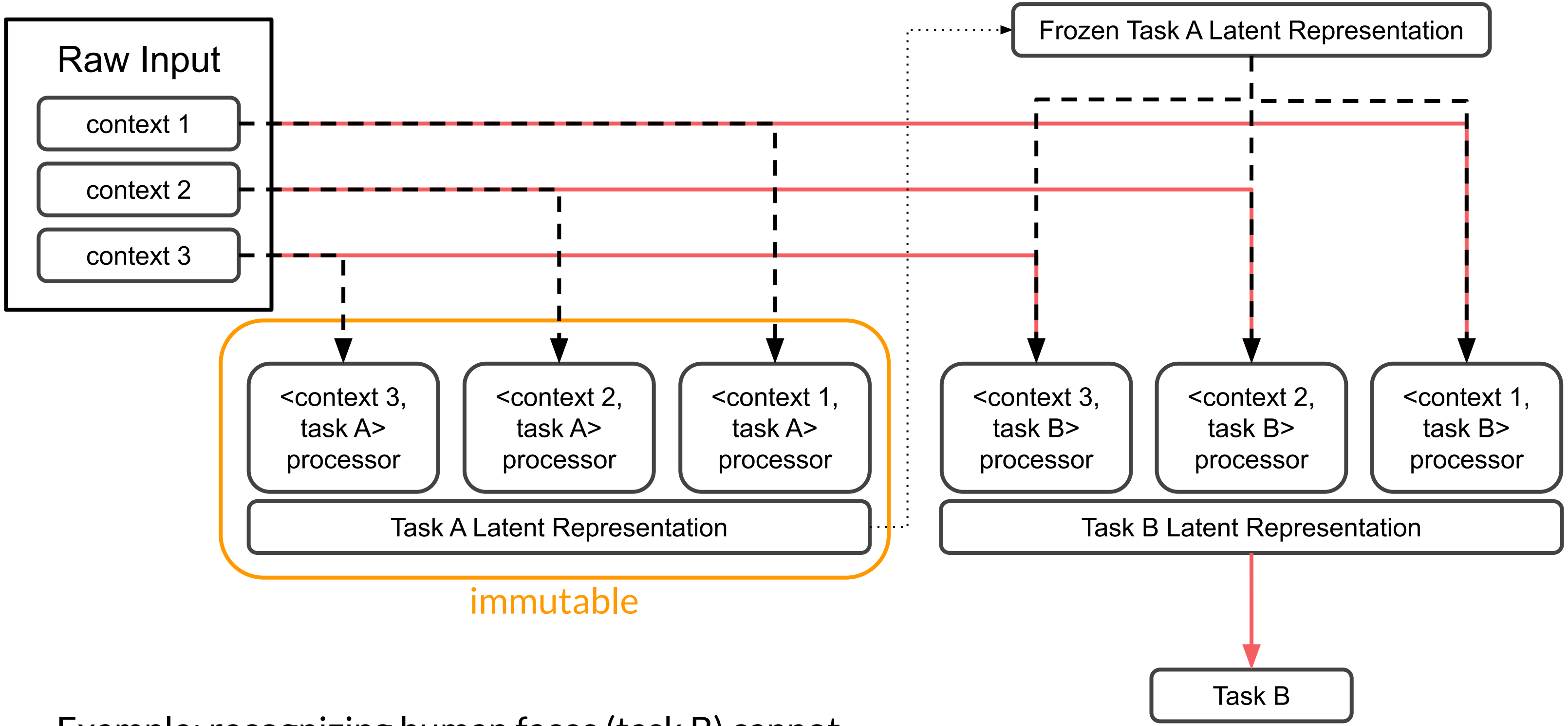
For HumanMeaning(TaskA) = Meaning(Task A Latent Representation) to hold true, task A's classes must be accurately predicted* across many contexts.
If there aren't enough contexts, there is no guarantee that Meaning(Task A Latent Representation) aligns with HumanMeaning(TaskA).

*prediction of classes or numerical values, or rewards from motor goals.

Two-task scenario

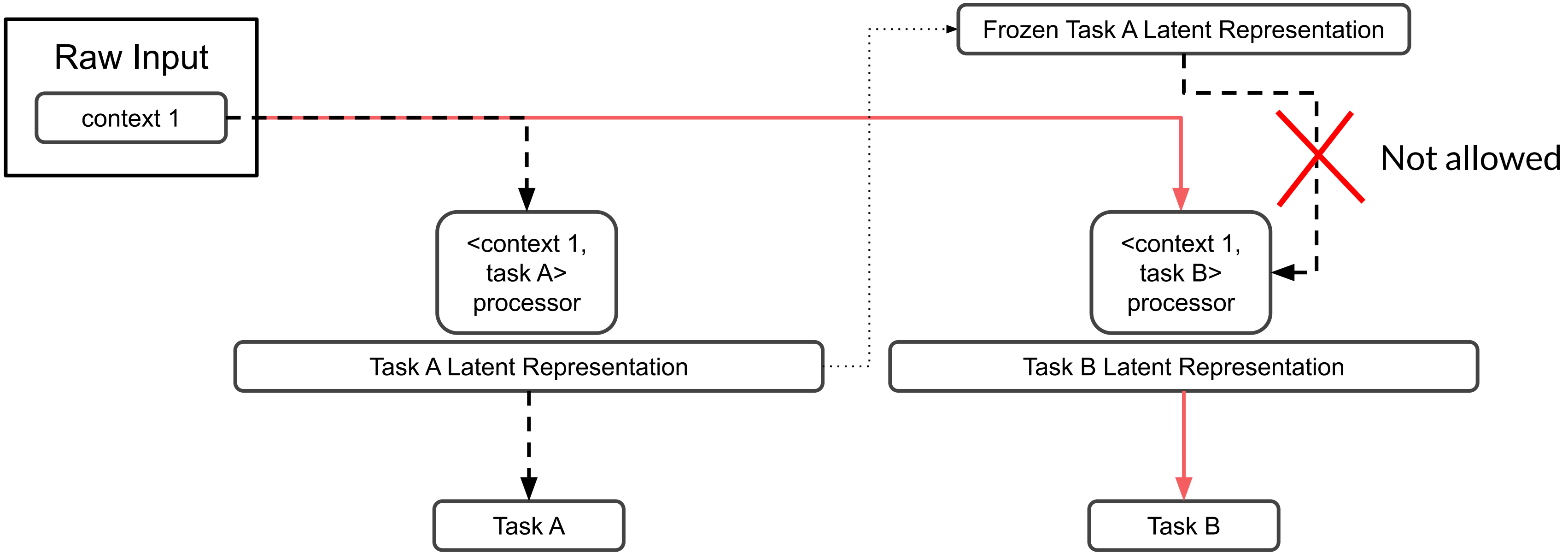

Frozen Task A Latent Representation
Raw Input
context 1
context 2
context 3
<context 3, task A> processor
<context 2, task A> processor
<context 1, task A> processor
<context 3, task B> processor
<context 2, task B> processor
<context 1, task B> processor
Task A Latent Representation
Task B Latent Representation
task A’s flow
Task A
Task B
task B’s flow

Rule 1: task A is frozen when training task B


Raw Input
context 1
context 2
context 3
<context 3, task A> processor
<context 2, task A> processor
<context 1, task A> processor
Task A Latent Representation
immutable
Frozen Task A Latent Representation
<context 3, task B> processor
<context 2, task B> processor
<context 1, task B> processor
Task B Latent Representation
Task B

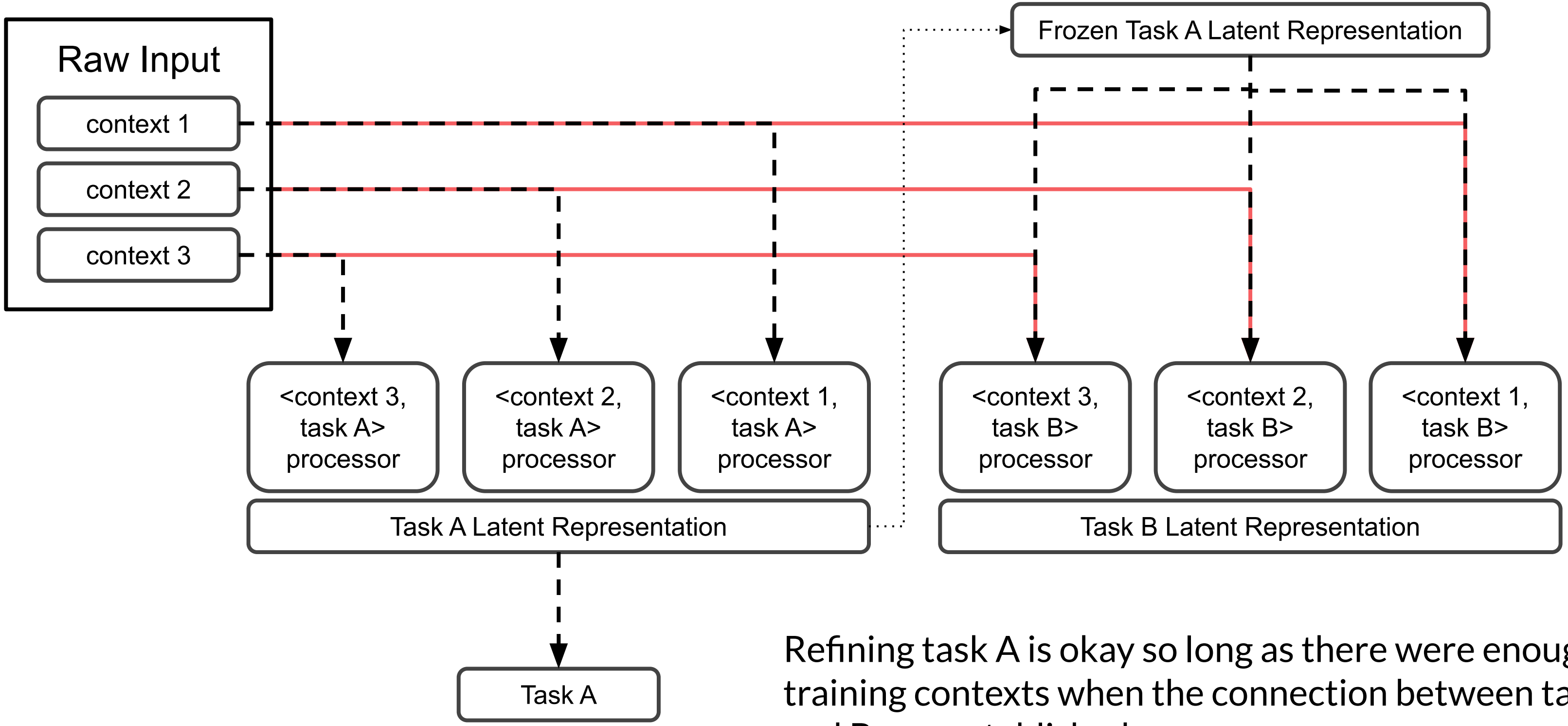
Example: recognizing human faces (task B) cannot interfere with the task of telling cats and dogs apart (task A).

Rule 2: task B cannot utilize task A if training contexts were scarce


Raw Input
context 1
Frozen Task A Latent Representation
Not allowed
<context 1, task A> processor
<context 1, task B> processor
Task A Latent Representation
Task B Latent Representation
Task A
Task B

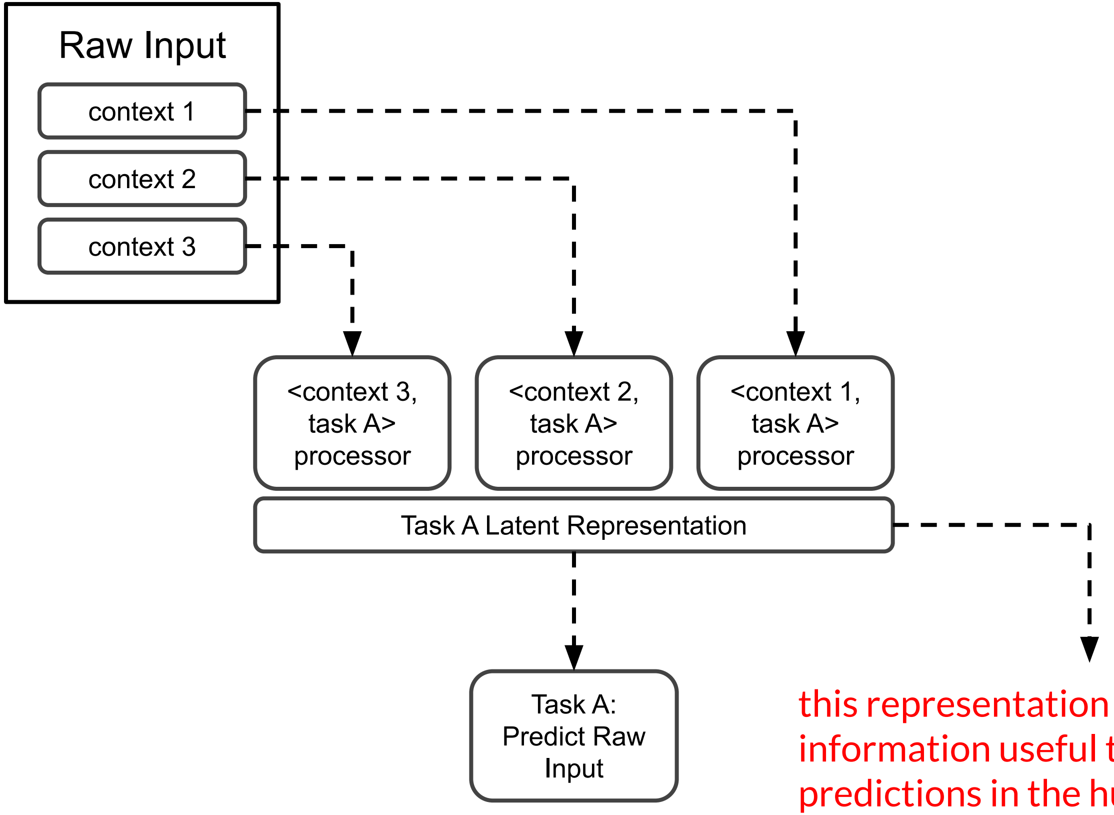
If there is a misalignment because of the lack of training contexts, i.e. HumanMeaning(TaskA) != Meaning(Task A LatentRepresentation),
then the system might find a correlation between Task A and Task B that doesn't exist in human reality.
Example: without this rule, the system might mistakenly connect cats (task A) to arctic foxes (task B) if white cats were the only kind of cats seen by the system. If a connection between task A and B were to be drawn, nothing would stop task A from interfering with task B in a destructive way.

Rule 3: Task A is allowed to interfere on task B under rule 2’s constraints


Frozen Task A Latent Representation
Raw Input
context 1
context 2
context 3
<context 3, task A> processor
<context 2, task A> processor
<context 1, task A> processor
<context 3, task B> processor
<context 2, task B> processor
<context 1, task B> processor
Task A Latent Representation
Task B Latent Representation
Task A

Refining task A is okay so long as there were enough training contexts when the connection between task A and B was established.

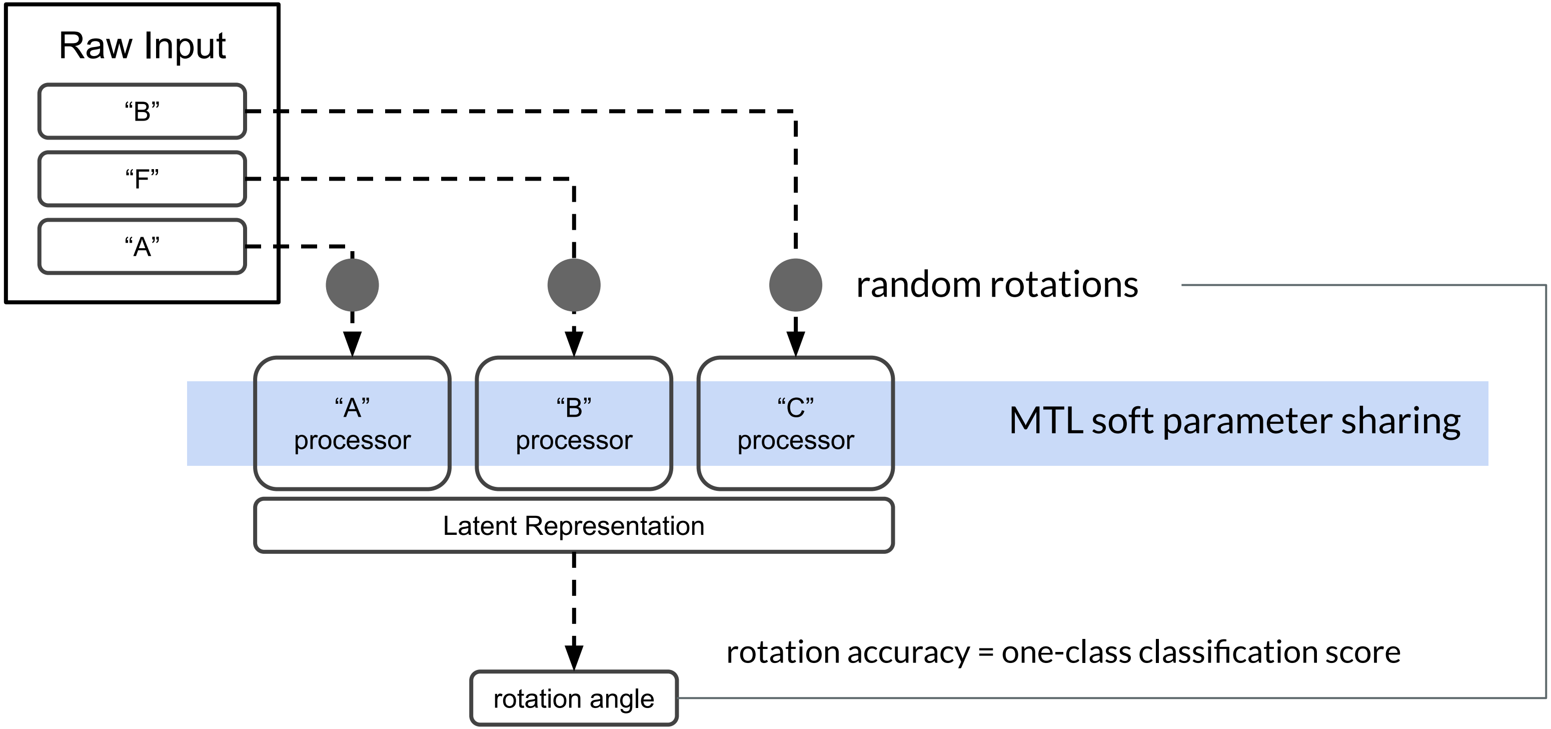
Rule 4: Tasks must interpret the raw input in subjective human terms


Raw Input
context 1
context 2
context 3
<context 3, task A> processor
<context 2, task A> processor
<context 1, task A> processor
Task A Latent Representation
Task A: Predict Raw Input

Autoencoding is an example of useless task.

this representation doesn’t carry information useful to making predictions in the human world more than the raw input itself.

One-task scenario: EMNIST



This works because “rotation” is a subjective concept that has the same meaning irrespective of the letter.
It is hard to guess the rotation angle without customizing the process for each letter.

Example of invalid task: if the task is to denoise letter images, it can be done without knowing the letter, thus it doesn’t help us predict the letter class.