

WHAT IS CLAIMED IS:

1. A computer-implemented method comprising:
obtaining, with at least one processor, training data; and
training, with at least one processor, a residual neural network including a first fully connected layer, a first recurrent neural network layer, and at least one skip connection, wherein the at least one skip connection directly connects at least one of (i) an output of the first fully connected layer to a first other layer downstream of the first recurrent neural network layer in the residual neural network and (ii) an output of the first recurrent neural network layer to a second other layer downstream of a second recurrent neural network layer in the residual neural network, by:
processing, with the first fully connected layer, the training data;
processing, with the first recurrent neural network layer, the output of the first fully connected layer;
at least one of: (i) processing, with the first other layer, the output of the first fully connected layer received via the at least one skip connection and an output of a first preceding layer between the first fully connected layer and the first other layer in the residual neural network and (ii) processing, with the second other layer, the output of the first recurrent neural network layer received via the at least one skip connection and an output of a second preceding layer between the first recurrent neural network layer and the second other layer in the residual neural network, wherein the residual neural network produces output data based on at least one of (i) an output of the first other layer and (ii) an output of the second other layer; and
modifying, using a loss function of the residual neural network that depends on the output data, one or more parameters of the residual neural network.
2. The computer-implemented method of claim 1, wherein the training data includes categorical data, numerical data, and embedding data, wherein the residual neural network further includes a categorical input layer, a numerical input layer, an embedding input layer, and a concatenation layer, and wherein training the residual neural network includes:

processing, with the categorical input layer, the categorical data to produce categorical features;

processing, with the numerical input layer, the numerical data to produce numerical features;

processing, with the embedding input layer, the embedding data to produce embedding features;

processing, with the concatenation layer, the categorical features, the numerical features, and the embedding features to produce a concatenated feature vector; and

processing, with the first fully connected layer, the concatenated feature vector.

3. The computer-implemented method of claim 1, wherein the residual neural network further includes a dropout layer downstream of the first recurrent neural network layer in the residual neural network, and wherein training the residual neural network includes:

processing, with the dropout layer, the output of the first recurrent neural network layer.

4. The computer-implemented method of claim 1, wherein the residual neural network further includes another fully connected layer as an output layer downstream of each other layer in the residual neural network, and wherein training the residual neural network includes:

processing, with the output layer, an output of a penultimate layer immediately upstream of the output layer in the residual neural network to produce the output data, wherein the output data includes a predicted label, wherein the predicted label includes a probability of the training data being associated with one or more predetermined classifications, and wherein the loss function of the residual neural network depends on the predicted label and an actual label associated with the training data.

5. The computer-implemented method of claim 1, wherein the at least one skip connection includes (i) a first skip connection that directly connects the output of the first fully connected layer to the first other layer downstream of the first

recurrent neural network layer in the residual neural network and (ii) a second skip connection that directly connects the output of the first recurrent neural network layer to the second other layer downstream of the second recurrent neural network layer in the residual neural network, wherein the first other layer includes the second recurrent neural network layer, wherein the second other layer includes a second fully connected layer, wherein the first preceding layer includes a first dropout layer between the first recurrent neural network layer and the second recurrent neural network layer, wherein the second preceding layer includes a second dropout layer between the second recurrent neural network layer and the second fully connected layer, and wherein training the residual neural network further includes:

processing, with the first dropout layer, the output of the first recurrent neural network layer;

processing, with the second recurrent neural network layer, an output of the first dropout layer and the output of the first fully connected layer received via the first skip connection;

processing, with the second dropout layer, an output of the second recurrent neural network layer; and

processing, with the second fully connected layer, an output of the second dropout layer and the output of the first recurrent neural network layer received via the second skip connection, wherein the residual neural network produces the output data based on an output of the second fully connected layer.

6. The computer-implemented method of claim 1, wherein the at least one skip connection directly connects the output of the first fully connected layer to the first other layer downstream of the first recurrent neural network layer in the residual neural network, wherein the first other layer includes a second fully connected layer, wherein the residual neural network further includes a first dropout layer between the first recurrent neural network layer and the second recurrent neural network layer, wherein the first preceding layer includes a second dropout layer between the second recurrent neural network layer and the second fully connected layer, and wherein training the residual neural network further includes:

processing, with the first dropout layer, the output of the first recurrent neural network layer;

processing, with the second recurrent neural network layer, an output of the first dropout layer;

processing, with the second dropout layer, an output of the second recurrent neural network layer; and

processing, with the second fully connected layer, an output of the second dropout layer and the output of the first fully connected layer received via the at least one skip connection, wherein the residual neural network produces the output data based on an output of the second fully connected layer.

7. The computer-implemented method of claim 1, wherein the at least one skip connection directly connects the output of the first fully connected layer to the first other layer downstream of the first recurrent neural network layer in the residual neural network, wherein the first other layer includes the second recurrent neural network, wherein the first preceding layer includes a first dropout layer, wherein the residual neural network further includes a second fully connected layer downstream of the second recurrent neural network and a second dropout layer between the second recurrent neural network and the second fully connected layer, and wherein training the residual neural network further includes:

processing, with the first dropout layer, the output of the first recurrent neural network layer;

processing, with the second recurrent neural network layer, an output of the first dropout layer and the output of the first fully connected layer received via the at least one skip connection;

processing, with the second dropout layer, an output of the second recurrent neural network layer; and

processing, with the second fully connected layer, an output of the second dropout layer, wherein the residual neural network produces the output data based on an output of the second fully connected layer.

8. The computer-implemented method of claim 1, wherein the at least one skip connection directly connects the output of the first recurrent neural network layer to the second other layer downstream of the second recurrent neural network layer in the residual neural network, wherein the second other layer includes a second fully connected layer, wherein the residual neural network further includes a

first dropout layer between the first recurrent neural network and the second recurrent neural network, wherein the second preceding layer includes a second dropout layer, and wherein training the residual neural network further includes:

- processing, with the first dropout layer, the output of the first recurrent neural network layer;

- processing, with the second recurrent neural network layer, an output of the first dropout layer;

- processing, with the second dropout layer, an output of the second recurrent neural network layer; and

- processing, with the second fully connected layer, an output of the second dropout layer and the output of the first recurrent neural network layer received via the at least one skip connection, wherein the residual neural network produces the output data based on an output of the second fully connected layer.

9. The computer-implemented method of claim 1, wherein the at least one skip connection directly connects the output of the first fully connected layer to the first other layer downstream of the first recurrent neural network layer in the residual neural network, wherein the first other layer includes a second fully connected layer, wherein the residual neural network further includes a first dropout layer between the first recurrent neural network layer and the second fully connected layer, wherein first recurrent neural network layer is the only recurrent neural network in the residual neural network, and wherein training the residual neural network further includes:

- processing, with the first dropout layer, the output of the first recurrent neural network layer; and

- processing, with the second fully connected layer, an output of the first dropout layer and the output of the first fully connected layer received via the at least one skip connection, wherein the residual neural network produces the output data based on an output of the second fully connected layer.

10. The computer-implemented method of claim 1, further comprising:

- providing, with at least one processor, the trained residual neural network;

obtaining, with at least one processor, input data; and
processing, with at least one processor, using the trained residual neural network, the input data to generate output data.

11. The computer-implemented method of claim 9, wherein the input data includes a request to authorize a transaction between a merchant and an account holder on an account issued by an issuer, and wherein the output data includes a prediction of whether the transaction is a fraudulent transaction.

12. A computing system comprising:
one or more processors programmed and/or configured to:
obtain training data; and
train a residual neural network including a first fully connected layer, a first recurrent neural network layer, and at least one skip connection, wherein the at least one skip connection directly connects at least one of (i) an output of the first fully connected layer to a first other layer downstream of the first recurrent neural network layer in the residual neural network and (ii) an output of the first recurrent neural network layer to a second other layer downstream of a second recurrent neural network layer in the residual neural network, by:

processing, with the first fully connected layer, the training data;

processing, with the first recurrent neural network layer, the output of the first fully connected layer;

at least one of: (i) processing, with the first other layer, the output of the first fully connected layer received via the at least one skip connection and an output of a first preceding layer between the first fully connected layer and the first other layer in the residual neural network and (ii) processing, with the second other layer, the output of the first recurrent neural network layer received via the at least one skip connection and an output of a second preceding layer between the first recurrent neural network layer and the second other layer in the residual neural network, wherein the residual neural network produces output

data based on at least one of (i) an output of the first other layer and (ii) an output of the second other layer; and

modifying, using a loss function of the residual neural network that depends on the output data, one or more parameters of the residual neural network.

13. The computing system of claim 12, wherein the training data includes categorical data, numerical data, and embedding data, wherein the residual neural network further includes a categorical input layer, a numerical input layer, an embedding input layer, and a concatenation layer, and wherein the one or more processors are further programmed and/or configured to the residual neural network by:

processing, with the categorical input layer, the categorical data to produce categorical features;

processing, with the numerical input layer, the numerical data to produce numerical features;

processing, with the embedding input layer, the embedding data to produce embedding features;

processing, with the concatenation layer, the categorical features, the numerical features, and the embedding features to produce a concatenated feature vector; and

processing, with the first fully connected layer, the concatenated feature vector.

14. The computing system of claim 12, wherein the residual neural network further includes a dropout layer downstream of the first recurrent neural network layer in the residual neural network, and wherein the one or more processors are further programmed and/or configured to the residual neural network by:

processing, with the dropout layer, the output of the first recurrent neural network layer.

15. The computing system of claim 12, wherein the residual neural network further includes another fully connected layer as an output layer downstream of each other layer in the residual neural network, and wherein the one or more

processors are further programmed and/or configured to the residual neural network by:

processing, with the output layer, an output of a penultimate layer immediately upstream of the output layer in the residual neural network to produce the output data, wherein the output data includes a predicted label, wherein the predicted label includes a probability of the training data being associated with one or more predetermined classifications, and wherein the loss function of the residual neural network depends on the predicted label and an actual label associated with the training data.

16. The computing system of claim 12, wherein the at least one skip connection includes (i) a first skip connection that directly connects the output of the first fully connected layer to the first other layer downstream of the first recurrent neural network layer in the residual neural network and (ii) a second skip connection that directly connects the output of the first recurrent neural network layer to the second other layer downstream of the second recurrent neural network layer in the residual neural network, wherein the first other layer includes the second recurrent neural network layer, wherein the second other layer includes a second fully connected layer, wherein the first preceding layer includes a first dropout layer between the first recurrent neural network layer and the second recurrent neural network layer, wherein the second preceding layer includes a second dropout layer between the second recurrent neural network layer and the second fully connected layer, and herein the one or more processors are further programmed and/or configured to the residual neural network by:

processing, with the first dropout layer, the output of the first recurrent neural network layer;

processing, with the second recurrent neural network layer, an output of the first dropout layer and the output of the first fully connected layer received via the first skip connection;

processing, with the second dropout layer, an output of the second recurrent neural network layer; and

processing, with the second fully connected layer, an output of the second dropout layer and the output of the first recurrent neural network layer received

via the second skip connection, wherein the residual neural network produces the output data based on an output of the second fully connected layer.

17. A computer program product comprising at least one non-transitory computer-readable medium including program instructions that, when executed by at least one processor, cause the at least one processor to:

obtain training data; and

train a residual neural network including a first fully connected layer, a first recurrent neural network layer, and at least one skip connection, wherein the at least one skip connection directly connects at least one of (i) an output of the first fully connected layer to a first other layer downstream of the first recurrent neural network layer in the residual neural network and (ii) an output of the first recurrent neural network layer to a second other layer downstream of a second recurrent neural network layer in the residual neural network, by:

processing, with the first fully connected layer, the training data;

processing, with the first recurrent neural network layer, the output of the first fully connected layer;

at least one of: (i) processing, with the first other layer, the output of the first fully connected layer received via the at least one skip connection and an output of a first preceding layer between the first fully connected layer and the first other layer in the residual neural network and (ii) processing, with the second other layer, the output of the first recurrent neural network layer received via the at least one skip connection and an output of a second preceding layer between the first recurrent neural network layer and the second other layer in the residual neural network, wherein the residual neural network produces output data based on at least one of (i) an output of the first other layer and (ii) an output of the second other layer; and

modifying, using a loss function of the residual neural network that depends on the output data, one or more parameters of the residual neural network.

18. The computer program product of claim 17, wherein the training data includes categorical data, numerical data, and embedding data, wherein the residual neural network further includes a categorical input layer, a numerical input

layer, an embedding input layer, and a concatenation layer, and wherein the instructions further cause the at least one processor to train the residual neural network by:

processing, with the categorical input layer, the categorical data to produce categorical features;

processing, with the numerical input layer, the numerical data to produce numerical features;

processing, with the embedding input layer, the embedding data to produce embedding features;

processing, with the concatenation layer, the categorical features, the numerical features, and the embedding features to produce a concatenated feature vector; and

processing, with the first fully connected layer, the concatenated feature vector.

19. The computer program product of claim 17, wherein the residual neural network further includes a dropout layer downstream of the first recurrent neural network layer in the residual neural network, and wherein the instructions further cause the at least one processor to train the residual neural network by:

processing, with the dropout layer, the output of the first recurrent neural network layer.

20. The computer program product of claim 17, wherein the residual neural network further includes another fully connected layer as an output layer downstream of each other layer in the residual neural network, and wherein the instructions further cause the at least one processor to train the residual neural network by:

processing, with the output layer, an output of a penultimate layer immediately upstream of the output layer in the residual neural network to produce the output data, wherein the output data includes a predicted label, wherein the predicted label includes a probability of the training data being associated with one or more predetermined classifications, and wherein the loss function of the residual neural network depends on the predicted label and an actual label associated with the training data.